



Universidad Politécnica de Madrid
Escuela Técnica Superior de Ingenieros Informáticos



POLITÉCNICA
"Ingeniamos el futuro"

Grupo de Investigación en Ingeniería del Software Empírica

Trabajo Fin de Máster
Máster Universitario en Software y Sistemas

**PERCEPCIONES SOBRE TRES TÉCNICAS DE EVALUACIÓN DE CÓDIGO:
UN ESTUDIO EMPÍRICO SOBRE LA INTUICIÓN Y LAS OPINIONES DE LOS
ENCARGADOS DE LAS PRUEBAS DE SOFTWARE**

Autor:
Patricia Riofrío Ojeda

Directores:
Sira Vegas
Natalia Juristo

Índice general

CAPÍTULO 1. INTRODUCCIÓN	1
1.1. Problema	1
1.2. Objetivos	2
1.3. Metodología de Trabajo.....	2
1.4. Estructura del Documento	3
CAPÍTULO 2. DESCRIPCIÓN DEL ESTUDIO EMPÍRICO	5
2.1. Objetivo, Preguntas de Investigación e Hipótesis	5
2.2. Planificación del Estudio Empírico	6
2.2.1. Contexto.....	6
2.2.2. Factores.....	6
2.2.3. Variables Respuesta	8
2.3. Diseño Experimental	9
2.3.1. Tipo de Diseño	9
2.3.2. Operación Experimental.....	10
2.3.3. Amenazas a la validez	10
2.4. Análisis de los Datos	10
2.4.1. Estudio de Intuiciones y Preferencias.....	12
2.4.2. Medida de Acuerdo.....	13
2.4.3. Direccionalidad del desacuerdo.....	15
2.4.4. Homogeneidad Marginal.....	16
2.4.5. Relación entre lo respondido y la realidad.....	16
CAPÍTULO 3. RESULTADOS SOBRE INTUICIÓN	19
3.1. Intuición sobre las Técnicas.....	19
3.1.1. ¿Qué técnica detecta más faltas?	19
3.1.2. ¿Qué técnica has aplicado mejor?.....	27
3.2. Intuición sobre los Programas.....	29
3.2.1. ¿Con qué programa has detectado más faltas?.....	29
3.2.2. ¿Con qué programa has detectado menos faltas?.....	31
3.3. Consistencia en las Contestaciones sobre las Técnicas.....	32
3.3.1. ¿Qué técnica detecta más faltas? / ¿Qué técnica has aplicado mejor?.....	33
3.4. Consistencia en las Contestaciones sobre los Programas.....	35
3.4.1. ¿Con qué programa has detectado más faltas? / ¿Con qué programa has detectado menos faltas?	35
CAPÍTULO 4. COMPARANDO LAS INTUICIONES CON LA REALIDAD	37
4.1. Intuiciones sobre las Técnicas.....	37
4.1.1. ¿Qué técnica detecta más faltas?	37
4.1.2. ¿Qué técnica has aplicado mejor?.....	52
4.2. Intuiciones sobre los Programas.....	63
4.2.1. ¿Con qué programa has detectado más faltas?.....	63

4.2.2. ¿Con qué programa has detectado menos faltas?	70
CAPÍTULO 5. RESULTADOS SOBRE PREFERENCIAS.....	75
5.1. Preferencias sobre las Técnicas	75
5.1.1. ¿Qué técnica te gusta más?	75
5.1.2. ¿Qué técnica te ha parecido más fácil de entender?	83
5.1.3. ¿Qué técnica te ha parecido más fácil de aplicar?	85
5.2. Preferencias sobre los Programas	92
5.2.1. ¿Qué programa es más fácil?	92
5.2.2. ¿Qué programa tiene las faltas más difíciles?	94
5.2.3. ¿Qué programa has entendido mejor?	96
5.2.4. ¿Qué programa tiene las faltas más fáciles?	98
5.3. Consistencia en las Contestaciones sobre las Técnicas.....	99
5.3.1. ¿Qué técnica te gusta más? / ¿Qué Técnica te ha parecido más fácil de entender?	100
5.3.2. ¿Qué técnica te gusta más? / ¿Qué técnica te ha parecido más fácil de aplicar?	102
5.3.3. ¿Qué técnica te ha parecido más fácil de entender? / ¿Qué técnica te ha parecido más fácil de aplicar?	106
5.4. Consistencias en las Contestaciones sobre los Programas.....	109
5.4.1. ¿Qué programa es más fácil? / ¿Qué programa tiene las faltas más difíciles?	109
5.4.2. ¿Qué programa es más fácil? / ¿Qué programa has entendido mejor?	111
5.4.3. ¿Qué programa tiene las faltas más difíciles? / ¿Qué programa has entendido mejor?	113
5.4.4. ¿Qué programa es más fácil? / ¿Qué programa tiene las faltas más fáciles?	115
5.4.5. ¿Qué programa tiene las faltas más difíciles? / ¿Qué programa tiene las faltas más fáciles?	117
5.4.6. ¿Qué programa has entendido mejor? / ¿Qué programa tiene las faltas más fáciles?	118
CAPÍTULO 6. COMPARANDO LAS PREFERENCIAS CON LA REALIDAD	120
6.1. Preferencias sobre las Técnicas	120
6.1.1. ¿Qué técnica te gusta más?	120
6.1.2. ¿Qué técnica te ha parecido más fácil de entender?	140
6.1.3. ¿Qué técnica te ha parecido más fácil de aplicar?	149
6.2. Preferencias sobre los programas	168
6.2.1. ¿Qué programa es más fácil?	168
6.2.2. ¿Qué programa tiene las faltas más difíciles?	174
6.2.3. ¿Qué programa has entendido mejor?	183
6.2.4. ¿Qué programa tiene las faltas más fáciles?	188
CAPÍTULO 7. COMPARANDO LA INTUICIÓN CON LAS PREFERENCIAS	191
7.1. Preguntas relacionadas con las Técnicas.....	191
7.1.1. ¿Qué técnica detecta más faltas? / ¿Qué técnica te gusta más?	191
7.1.2. ¿Qué técnica detecta más faltas? / ¿Qué técnica te ha parecido más fácil de entender?	196
7.1.3. ¿Qué técnica detecta más faltas? / ¿Qué técnica te ha parecido más fácil de aplicar?	199
7.1.4. ¿Qué técnica has aplicado mejor? / ¿Qué técnica te gusta más?	203
7.1.5. ¿Qué técnica has aplicado mejor? / ¿Qué técnica te ha parecido más fácil de entender?	205
7.1.6. ¿Qué técnica has aplicado mejor? / ¿Qué técnica es más fácil de aplicar?	208
7.2. Preguntas relacionadas con los Programas.....	210
7.2.1. ¿Con qué programa has detectado más faltas? / ¿Qué programa es más fácil?	210
7.2.2. ¿Con qué programa has detectado más faltas? / ¿Qué programa tiene las faltas más difíciles?	212
7.2.3. ¿Con qué programa has detectado más faltas? / ¿Qué programa has entendido mejor?	214
7.2.4. ¿Con qué programa has detectado más faltas? / ¿Qué programa tiene las faltas más fáciles?	216
7.2.5. ¿Con qué programa has detectado menos faltas? / ¿Qué programa es más fácil?	217

7.2.6. ¿Con qué programa has detectado menos faltas? / ¿Qué programa tiene las faltas más difíciles?	219
7.2.7. ¿Con qué programa has detectado menos faltas? / ¿Qué programa has entendido mejor?	220
7.2.8. ¿Con qué programa has detectado menos faltas? / ¿Qué programa tiene las faltas más fáciles?..	222
CAPÍTULO 8. DISCUSIÓN	224
8.1. Resultados sobre Intuición	224
8.1.1. Resultados sobre las Técnicas	224
8.1.2. Resultados sobre los Programas	226
8.1.3. Consistencias entre las Preguntas	226
8.1.4. Resultados Globales	226
8.2. Resultados sobre Preferencias.....	227
8.2.1. Resultados sobre las Técnicas	227
8.2.2. Resultados sobre los Programas	230
8.2.3. Consistencias entre las Preguntas	230
8.2.4. Resultados Globales	232
8.3. Acuerdo entre las Intuiciones y Preferencias	232
8.3.1. Resultados sobre las Técnicas	233
8.3.2. Resultados sobre los Programas	234
8.3.3. Resultados Globales	236
8.4. Intuiciones vs Realidad	236
8.4.1. Resultados sobre las Técnicas	236
8.4.2. Resultados sobre los Programas	238
8.4.3. Resultados Globales	238
8.5. Preferencias vs Realidad	239
8.5.1. Resultados sobre las Técnicas	239
8.5.2. Resultados sobre los Programas	241
8.5.3. Resultados Globales	243
CAPÍTULO 9. CONCLUSIONES.....	245
REFERENCIAS.....	249

RESUMEN

La Ingeniería del Software Empírico (ISE) utiliza como herramientas los estudios empíricos para conseguir evidencias que ayuden a conocer bajo qué circunstancias es mejor usar una tecnología software en lugar de otra. La investigación en la que se enmarca este TFM explora si las intuiciones y/o preferencias de las personas que realizan las pruebas de software, son capaces de predecir la efectividad de tres técnicas de evaluación de código: lectura por abstracciones sucesivas, cobertura de decisión y partición en clases de equivalencia.

Para conseguir dicho objetivo, se analizan los datos recogidos en un estudio empírico, realizado por las tutoras de este TFM. En el estudio empírico distintos sujetos aplican las tres técnicas de evaluación de código a tres programas distintos, a los que se les habían introducido una serie de faltas artificialmente. Los sujetos deben reportar los fallos encontrados en los programas, así como, contestar a una serie de preguntas sobre sus intuiciones y preferencias.

A la hora de analizar los datos del estudio, se ha comprobado: 1) cuáles son sus intuiciones y preferencias (mediante el test estadístico X^2 de Pearson); 2) si los sujetos cambian de opinión después de aplicar las técnicas (para ello se ha utilizado índice de Kappa, el Test de McNemar-Bowker y el Test de Stuart-Maxwell); 3) la consistencia de las distintas preguntas (mediante el índice de Kappa), comparando: intuiciones con intuiciones, preferencias con preferencias e intuiciones con preferencias; 4) Por último, si hay coincidencia entre las intuiciones y preferencias con la efectividad real obtenida (para ello se ha utilizado, el Modelo Lineal General con medidas repetidas).

Los resultados muestran que, no hay una intuición clara ni tampoco una preferencia concreta, con respecto a los programas. Además aunque existen cambios de opinión después de aplicar las técnicas, no se encuentran evidencias claras para afirmar que la intuición y preferencias influyen en su efectividad. Finalmente, existen relaciones entre las intuiciones con intuiciones, preferencias con preferencias e intuiciones con preferencias, además esta relación es más notoria después de aplicar las técnicas.

ABSTRACT

Empirical Software Engineering (ESE) uses empirical studies as a mean to generate evidences to help determine under what circumstances it is convenient to use a given software technology. This Master Thesis is part of a research that explores whether intuitions and/or preferences of testers, can be used to predict the effectiveness of three code evaluation techniques: reading by stepwise abstractions, decision coverage and equivalence partitioning.

To achieve this goal, this Master Thesis analyzes the data collected in an empirical study run by the tutors. In the empirical study, different subjects apply three code evaluation techniques to three different programs. A series of faults were artificially introduced to the programs. Subjects are required to report the defects found in the programs, as well as answer a series of questions about their intuitions and preferences.

The data analyses test: 1) what are the intuitions and preferences of the subjects (using the Pearson X2 test); 2) whether subjects change their minds after applying the techniques (using the Kappa coefficient, McNemar-Bowker test, and Stuart-Maxwell test); 3) the consistency of the different questions, comparing: intuitions versus intuitions, preferences versus preferences and preferences versus intuitions (using the Kappa coefficient); 4) finally, if intuitions and/or preferences predict the actual effectiveness obtained (using the General Linear Model, repeated measures).

The results show that there is not clear intuition or particular preference with respect to the programs. Moreover, although there are changes of mind after applying the techniques, there are not clear evidences to claim that intuition and preferences influence their effectiveness. Finally, there is a relationship between the intuitions versus intuitions, preferences versus preferences and intuitions versus preferences; this relationship is more noticeable after applying the techniques.

CAPÍTULO 1. INTRODUCCIÓN

El campo de la Ingeniería del Software Empírica (ISE) centra su trabajo en la evaluación de las tecnologías utilizadas para el desarrollo del software. Esta disciplina busca evidencias que ayuden a afirmar bajo qué condiciones es mejor usar una tecnología en lugar de otra. Estas evidencias pueden ayudar a producir software de mejor calidad y con un menor consumo de recursos.

La ISE utiliza los estudios empíricos como medio para conseguir las evidencias anteriormente mencionadas. Dichas evidencias aportan una base teórica a las suposiciones y especulaciones que actualmente se tienen sobre las tecnologías utilizadas en el desarrollo del software. Esta metodología de trabajo ha sido heredada de otras disciplinas tales como la medicina y la psicología.

El ámbito en el que se desarrolla este TFM es un estudio empírico llevado a cabo en la Universidad Politécnica de Madrid, por el Grupo de Investigación en Ingeniería de Software Experimental de la misma universidad (en concreto, por las dos directoras de este TFM). El fin de dicho estudio empírico, es comparar la efectividad de tres técnicas de evaluación de código.

La investigación en la que se enmarca este TFM, pretende averiguar si las distintas preferencias e intuiciones, de las personas que realizaron las pruebas de software, coinciden con la efectividad obtenida al aplicar las técnicas de evaluación de código.

1.1. Problema

La intuición es una idea que se genera en el cerebro de manera inconsciente. Al hablar de intuición se hablan de *ideas sentidas* más que *pensadas*. Por lo tanto, ¿existe intuición con respecto a las técnicas de evaluación de código?

Por otro lado, existen dos tipos de intuición: la primera son los juicios intuitivos retrospectivos, y la segunda son las inferencias prospectivas o las predicciones. Por lo tanto, se quiere averiguar, si en el caso de que exista intuición, de cuál de los dos tipos sería la misma, es decir, se puede usar para predecir que una técnica será aplicada mejor que otra.

Finalmente, se conoce que cuando el ser humano está motivado es mucho más eficiente y productivo. Esta motivación se adquiere fácilmente cuando la actividad realizada es grata para el individuo. Por esta razón, se quiere analizar si las preferencias que se tienen sobre las técnicas de evaluación de código influyen en los sujetos para que una técnica sea aplicada mejor que otra.

1.2. Objetivos

El objetivo del presente TFM es realizar el análisis de los datos recogidos en el estudio empírico realizado por las directoras de este TFM, sobre las técnicas de evaluación de código¹. Por lo tanto, el objetivo general de este trabajo consiste en estudiar si las intuiciones y/o preferencias de los sujetos al de aplicar tres técnicas de evaluación de código coinciden con la efectividad de las mismas. Este objetivo general se ha descompuesto en los siguientes subobjetivos:

- Conocer las intuiciones y las preferencias de los sujetos en lo relativo a las técnicas de evaluación estudiadas.
- Averiguar si el hecho de aplicar las técnicas produce un cambio de opinión en los sujetos acerca de sus intuiciones y preferencias.
- Averiguar si las intuiciones y las preferencias sirven para predecir la efectividad de las técnicas.
- Estudiar si existe relación entre las intuiciones y las preferencias de los sujetos.

1.3. Metodología de Trabajo

Para el desarrollo de este trabajo, se realiza un estudio empírico cuyo objetivo es determinar la efectividad de tres técnicas de evaluación de código: una técnica estática, una técnica estructural y una técnica funcional, así como identificar las intuiciones y preferencias de los sujetos sobre las mismas.

El estudio empírico, se realiza en tres sesiones y en cada sesión los sujetos experimentales aplican una de las técnicas mencionadas, a un programa determinado. En una sesión los sujetos aplican distintas técnicas, de tal modo que en una sesión se consigue que se apliquen las tres técnicas.

Los sujetos reportan los defectos encontrados al aplicar las técnicas sobre los programas. Cada programa contiene una serie de faltas introducidas artificialmente. Asimismo, los sujetos rellenan unos cuestionarios con preguntas sobre sus intuiciones y preferencias. Estos cuestionarios se rellenan antes de aplicar las técnicas y también después de aplicar las mismas. Por lo tanto, hay preguntas que aparecen en los cuestionarios tanto antes como después de la aplicación de las técnicas.

¹ Nótese que este TFM forma parte de una investigación más amplia. Los estudios empíricos ya estaban hechos, se realizaron durante tres años consecutivos. En este TFM lo que se ha realizado son los análisis de los datos de dichos estudios.

Para poder alcanzar los objetivos mencionados anteriormente, se realizan los análisis estadísticos necesarios relativos a la intuición y preferencia sobre las técnicas y programas.

Lo primero, hay que analizar cuáles son las intuiciones y preferencias de los sujetos, para ello se utiliza el Test X^2 de Pearson. Por otro lado, también se analiza si los sujetos cambian de opinión al aplicar las técnicas, para ello se utiliza el índice de Kappa, el test de McNemar-Bowker y el test de Stuart-Maxwell.

Con el objetivo de comprobar la consistencia entre las preguntas se analiza el acuerdo existente entre las preguntas del mismo tipo entre sí (intuición frente a intuición y preferencias frente a preferencias) y también entre los dos tipos de preguntas (intuición frente a preferencias).

Finalmente, se comprueba si existe o no relación entre las intuiciones y preferencias de los sujetos con la realidad, para ello se utiliza, el Modelo Lineal General con medidas repetidas.

1.4. Estructura del Documento

A continuación se describe la organización del documento.

En el capítulo 2, se detalla el estudio empírico del cual se han obtenido los datos para poder realizar la investigación aquí descrita. Asimismo, se explican cada uno de los métodos estadísticos utilizados para el análisis de los datos. Este capítulo sirve de guía para entender los capítulos posteriores, puesto que a partir del capítulo 3 y hasta el 7, únicamente se comentarán los resultados obtenidos.

En el capítulo 3, se analizan todas las preguntas relacionadas con la intuición. Esto implica conocer las intuiciones de los sujetos relativas a las técnicas y programas, además comprobar si existen cambios de opinión en las mismas y finalmente, estudiar si hay relación entre las contestaciones a las distintas preguntas.

En el capítulo 4, se analiza la relación existente en las respuestas dadas con respecto a la intuición y los resultados obtenidos sobre efectividad en el estudio empírico, con el fin de determinar si existe relación entre las intuiciones y el comportamiento de las técnicas de evaluación de código.

En el capítulo 5, se analizan todas las preguntas relacionadas con las preferencias. Esto implica conocer las preferencias de los sujetos relativas a las técnicas y programas, además comprobar si existen cambios de opinión en las mismas y finalmente, estudiar si hay relación entre las distintas preguntas.

En el capítulo 6, se analiza la relación existente en las respuestas dadas con respecto a las preferencias y los resultados obtenidos sobre efectividad en el estudio empírico, con el fin de determinar si existe relación entre las preferencias y el comportamiento de las técnicas de evaluación de código.

En el capítulo 7, se analiza la posible relación entre las preguntas de intuición frente a las preguntas de preferencias, para comprobar si existe objetividad en las intuiciones de los sujetos, o por el contrario comprobar si las repuestas que han proporcionado se basan en sus preferencias.

En el capítulo 8, se presenta una discusión de los capítulos 3 hasta el 7, destacando los resultados relevantes obtenidos en el análisis de cada capítulo. En general, en este capítulo se presenta un resumen a alto nivel de los resultados obtenidos en el desarrollo de esta investigación.

Finalmente en el capítulo 9, se tratan las conclusiones del trabajo realizado.

CAPÍTULO 2. DESCRIPCIÓN DEL ESTUDIO EMPÍRICO

En este capítulo se describe el estudio empírico mediante el cual se han obtenido los datos para poder realizar el estudio planteado en esta investigación. Se siguen las ideas descritas en [1].

2.1. Objetivo, Preguntas de Investigación e Hipótesis

El objetivo de este trabajo consiste en estudiar si las intuiciones y/o preferencias de los sujetos al de aplicar tres técnicas de evaluación de código coinciden con la efectividad de las mismas. Este objetivo general se ha descompuesto en una serie de subobjetivos presentados en la sección 1.2. Del objetivo general y subobjetivos concretos se pueden derivar las siguientes preguntas de investigación (PI) e hipótesis (H):

PI 1: ¿Cuáles son las intuiciones y las preferencias de los sujetos en lo relativo a las técnicas de evaluación estudiadas?

- H_{10} : No existe una intuición clara de los sujetos por alguna de las técnicas.
- H_{11} : Existe una intuición clara de los sujetos por alguna de las técnicas.
- H_{20} : No existe una preferencia clara de los sujetos por alguna de las técnicas.
- H_{21} : Existe una preferencia clara de los sujetos por alguna de las técnicas.

PI 2: ¿El hecho de aplicar las técnicas produce un cambio de opinión en los sujetos acerca de sus intuiciones y preferencias?

- H_{30} : Una vez aplicadas las técnicas los sujetos no cambian sus intuiciones
- H_{31} : Una vez aplicadas las técnicas los sujetos cambian sus intuiciones

- H_{40} : Una vez aplicadas las técnicas los sujetos no cambian sus preferencias
- H_{41} : Una vez aplicadas las técnicas los sujetos cambian sus preferencias

PI 3: ¿Sirven las intuiciones y las preferencias para predecir la efectividad de las técnicas?

- H_{50} : La intuición no predice que una técnica sea más efectiva que otra
- H_{51} : La intuición predice que una técnica sea más efectiva que otra
- H_{60} : Las preferencias no predicen que una técnica sea más efectiva que otra
- H_{61} : Las preferencias predicen que una técnica sea más efectiva que otra.

PI 4: ¿Existe relación entre las intuiciones y las preferencias de los sujetos?

- H_{60} : Las intuiciones no están relacionadas con las preferencias
- H_{61} : Las intuiciones están relacionadas con las preferencias.

2.2. Planificación del Estudio Empírico

Este apartado, contiene la información sobre el contexto donde fue realizado el estudio empírico, así como el detalle de sus variables respuestas y factores.

2.2.1. Contexto

El estudio empírico se realizó, durante tres años consecutivos, con los alumnos de quinto curso de Ingeniería Informática de la Facultad de Informática de la Universidad Politécnica de Madrid (UPM).

El estudio empírico formaba parte de la asignatura “Evaluación de Sistemas de Información”. En dicha asignatura, se enseñaban conceptos y técnicas de verificación y validación de Software. Por lo tanto, en el momento de realizar el estudio empírico, los sujetos ya están familiarizados con las técnicas de evaluación.

2.2.2. Factores

Los factores considerados en el estudio son: las técnicas de evaluación de código y los programas usados.

2.2.2.1. Técnicas

Para el estudio del estudio empírico, se han aplicado tres técnicas distintas: una estática y dos dinámicas (una estructural y otra funcional). La técnica estática que se ha utilizado es *lectura por*

abstracciones sucesivas, la técnica estructural es *cobertura de decisión* y la técnica funcional es *partición en clases de equivalencia*. A continuación, se describen cada una de las técnicas utilizadas.

Para aplicar la técnica de *lectura por abstracciones sucesivas* [2] los sujetos comienzan en el nivel más bajo (más detallado) y reemplazan las líneas consecutivas de código del programa por una abstracción (o especificación) que resume los posibles resultados, con independencia de la estructura de control interno y las operaciones de datos. Los sujetos repiten el proceso hasta que hayan abstraído todo el código fuente. A continuación, los sujetos comparan las especificaciones del programa con su abstracción para observar las inconsistencias entre el comportamiento del programa específico y el esperado. Se ha proporcionado a los sujetos el código fuente, a partir del cual construirán una especificación del programa. A continuación, los sujetos comparan la especificación creada con la real. Cualquier diferencia entre las dos especificaciones se considera una falta.

La técnica de *cobertura de decisión* [2] tiene como objetivo garantizar que para cada uno de los puntos de decisión del código, se ejecuten todas las ramas al menos una vez. De esta manera se garantiza que se ejecuta todo el código. Al validar todas las ramas en el código, se consigue comprobar que ninguna rama conduzca a un comportamiento anormal del programa. Cada rama toma un trayecto verdadero o falso. Una decisión es una instrucción IF, una declaración de control de bucle (DO-WHILE o REPEAT-UNTIL), o una instrucción CASE, donde hay dos o más resultados a partir de la declaración. Con una instrucción IF, la salida puede ser verdadera o falsa, según el valor de la condición lógica que viene después de IF. Para ello se ha proporcionado a los sujetos el código fuente sin la especificación, para que realizaran los casos de prueba, como especifica la técnica, y obtener resultados. Los resultados obtenidos se comparan con la especificación inicial. Los resultados incorrectos representan las faltas del programa.

El procedimiento de la técnica *partición en clases de equivalencia* [2] se divide en dos pasos: primero hay que identificar las clases de equivalencia y luego se definen los casos de prueba. Las clases de equivalencia se identifican a través de las condiciones de entrada de un programa (normalmente una frase o declaración en la especificación). Se generan dos tipos de clases de equivalencia, las clases válidas, que representan los valores válidos para las entradas del programa, y las clases inválidas, que representan los valores erróneos para las entradas del programa. Las clases de equivalencia se identifican basándose en si en las condiciones de entrada se especifican un rango de valores, un número de valores, un conjunto de valores de entrada o una situación "tiene que ser". El segundo paso consiste en utilizar las clases de equivalencia para identificar los casos de prueba. El objetivo es probar todas las clases válidas e inválidas al menos una vez. En un caso de prueba pueden aparecer varias clases válidas, sin embargo, para las clases inválidas solo pueden aparecer una en cada caso de prueba, con el fin de evitar que los errores se enmascaren mutuamente. Por lo tanto, se ha proporcionado a los sujetos la versión ejecutable del programa y la especificación asociada al mismo.

2.2.2.2. Programas

Para el estudio se utilizaron tres programas. Los programas están escritos en el lenguaje de programación C y contienen una media de 200 líneas de código, excluyendo líneas en blanco y comentarios. A continuación se explica brevemente cada uno:

- Cmdline: Programa que recibe como entrada una línea, la analiza y devuelve un resumen del análisis de la misma. Tiene 209 líneas de código y una complejidad ciclomática de 61.
- Nametbl: Programa que implementa la estructura de datos y las operaciones de una tabla de símbolos. Tiene 172 líneas de código y una complejidad ciclomática de 29.
- Ntree: Programa que implementa la estructura de datos y las operaciones de un árbol n-ario. Tiene 146 líneas de código y una complejidad ciclomática de 21.

A cada programa, se le han insertado artificialmente una serie de faltas. En la primera iteración cada programa tiene 7 faltas, mientras que en la segunda y tercera iteración cada uno tiene 6 faltas. Los conjuntos de faltas de los programas en la primera iteración son distintos a los de la segunda y tercera iteración. Esto es debido, a que, en la primera iteración todas las faltas son detectables por todas las técnicas, sin embargo, en las iteraciones 2 y 3 todas las faltas son detectables solo por las técnicas funcional y estructural.

2.2.3. Variables Respuesta

Durante la realización del estudio empírico se obtuvieron dos tipos de variables respuestas: cuantitativas y cualitativas.

- Los datos *cuantitativos* son la efectividad de las técnicas medida como el porcentaje de faltas encontradas por los sujetos en un determinado programa al aplicar una técnica sobre él.
- Los datos *cualitativos* son las respuestas a unos cuestionarios sobre intuiciones y preferencias que los sujetos contestaron antes y después de aplicar las técnicas.

La Tabla 2.1 y la Tabla 2.2 muestran las preguntas realizadas. Estas se han dividido en preguntas sobre intuición y preguntas sobre preferencias, respectivamente. En cada tabla, se muestra un código que se ha asignado a cada pregunta, para facilitar la identificación de la misma, donde la primera letra indica si la pregunta está relacionada con las intuiciones o las preferencias (I ó P), la segunda letra indica si se trata de preguntas sobre las técnicas o sobre los programas (T ó P), finalmente se han enumerado todas (1, 2, ... n), primero las preguntas relacionadas con las técnicas y luego para las relacionadas con los programas se vuelve a empezar en 1. Finalmente, en la misma tabla se puede observar la iteración, es decir estudio empírico, en la que aparece cada una de las preguntas y si fueron realizadas antes o después de aplicar las técnicas de evaluación de código. Además se puede observar que se han dividido en cada tabla, las preguntas relacionadas con las técnicas y las relacionadas con los programas.

Código	Preguntas sobre Intuición	It 1	It 2	It 3	Aparición
IT1	¿Qué técnica detecta más faltas?	Si	Si	Si	A/D
IT2	¿Qué técnica has aplicado mejor?	Si	Si	Si	D
IP1	¿Con qué programa has detectado más faltas?	No	Si	Si	D
IP2	¿Con qué programa has detectado menos faltas?	No	No	Si	D

Tabla 2.1. Preguntas sobre Intuición

Código	Preguntas sobre Preferencias	It 1	It 2	It 3	Aparición
PT1	¿Qué técnica te gusta más?	Si	Si	Si	A/D
PT2	¿Qué técnica te ha parecido más fácil de entender?	Si	Si	Si	D
PT3	¿Qué técnica te ha parecido más fácil de aplicar?	Si	Si	Si	A/D
PP1	¿Qué programa es más fácil?	Si	Si	Si	D
PP2	¿Qué programa tiene las faltas más difíciles?	Si	Si	Si	D
PP3	¿Qué programa has entendido mejor?	No	Si	Si	D
PP1	¿Qué programa tiene las faltas más fáciles?	No	No	Si	D

Tabla 2.2. Preguntas sobre Preferencia

2.3. Diseño Experimental

En este apartado se describe en diseño experimental seguido en el estudio empírico.

2.3.1. Tipo de Diseño

Se ha utilizado un diseño de medidas repetidas tipo cross-over. En este tipo de diseños los sujetos aplican todos los tratamientos, obteniéndose de este modo tantas medidas para cada variable respuesta por cada sujeto, como tratamientos existan.

En este diseño hay tres tratamientos para el factor técnica (revisión, estructural y funcional) y tres tratamientos para el factor programa (cmdline, nametbl y ntree). Cada sujeto aplica una técnica distinta sobre un programa. Asimismo, no todos los sujetos aplican las mismas técnicas sobre los mismos programas.

El estudio empírico se realizó con el fin de evaluar la efectividad de tres técnicas de evaluación de código. El total de sujetos experimentales (29, 46, 46, respectivamente en cada iteración) se dividió en grupos de tamaño uniforme. Cada día todos los sujetos aplicaban todas las técnicas sobre el mismo programa, con el fin de evitar que en los días sucesivos los sujetos tuvieran conocimientos previos de la especificación de los programas, contemplando así la posibilidad de que se hubiesen comentado entre ellos los programas y sus fallos. La tabla 2.3 muestra el diseño del experimento.

	Día 1			Día 2			Día 3		
Técnica	R	E	F	R	E	F	R	E	F
Grupo 1	X	-	-	-	X	-	-	-	X
Grupo 2	X	-	-	-	-	X	-	X	-
Grupo 3	-	X	-	-	-	X	X	-	-
Grupo 4	-	X	-	X	-	-	-	-	X
Grupo 5	-	-	X	X	-	-	-	X	-
Grupo 6	-	-	X	-	X	-	X	-	-

Tabla 2.3. Diseño del Estudio empírico

2.3.2. Operación Experimental

El procedimiento experimental consta de tres sesiones. En cada sesión, los distintos sujetos, aplican todas las técnicas sobre un único programa. De esta manera, el orden de aplicación de las técnicas: (Funcional-Estructural-Revisión, Funcional-Revisión-Estructural, Estructural-Revisión-Funcional, Estructural-Funcional-Revisión, Revisión-Funcional-Estructural y Revisión-Estructural-Funcional) está cubierto a través de las tres sesiones.

En cada sesión se aplican las tres técnicas, por lo tanto, las tres técnicas se aplican sobre el mismo programa. Esto lleva a confundir el día de aplicación de la técnica con el programa, por lo tanto, los programas se aplican en diferente orden en cada iteración. El orden de aplicación de los programas, se puede observar en la Tabla 2.4.

Cada sesión tiene una duración de cuatro horas, lo que equivale a que no hay límite de tiempo, ya que el tiempo es suficiente para completar una tarea. Los sujetos reciben tres sesiones de entrenamiento de cuatro horas para aprender a aplicar las técnicas antes del estudio empírico.

	Día 1	Día 2	Día 3
Iteración 1	ntree	cmdline	nametbl
Iteración 2	cmdline	ntree	nametbl
Iteración 3	ntree	nametbl	cmdline

Tabla 2.4. Orden de aplicación de los programas

2.3.3. Amenazas a la validez

El estudio tiene algunas amenazas a la validez que indican que hay que tener cuidado sobre cómo son usados los resultados [3], ya que:

- Los resultados se han obtenido a partir de testers novatos. Si los sujetos fueran expertos, los resultados podrían ser diferentes.
- Los resultados se han obtenido a partir de tres programas escritos en C. Si se utilizan programas más grandes, o escritos en un lenguaje de programación diferente, los resultados podrían ser diferentes.
- Los resultados se han obtenido de una precisa aplicación de las técnicas. Por ejemplo, un analizador no dinámico se usa para aplicar las técnicas de cobertura de decisión.
- Las variables identificadas durante la reunión con los investigadores, que ejecutaron los experimentos, podrían influir en los resultados obtenidos. Dichas variables se deberán analizar más en futuros experimentos para comprobar si hay influencia.

2.4. Análisis de los Datos

A continuación se describe el análisis realizado, de modo general, así como cada uno de los test estadísticos concretos que se aplican para el análisis de los datos. Para el análisis de los datos se utiliza el programa estadístico IBM SPSS Statistics Versión 20.

Como ya se ha comentado en la sección 2.2.3, existen tres variables respuesta: intuición, preferencias y el porcentaje de faltas encontradas. Por lo tanto, primero, se analizan las intuiciones,

luego las preferencias, seguidamente se comparan las intuiciones frente a las preferencias y finalmente, se comparan las intuiciones y preferencias frente a la efectividad real. En concreto:

- **Estudio de intuiciones y preferencias.** Con respecto a cada pregunta, primero es necesario estudiar la distribución de la respuesta, para tener una idea clara de que es lo que responden los sujetos. Para esto se van a utilizar gráficos circulares y el test estadístico Chi-Cuadrado de Pearson [4]. En el capítulo 3, se analizan la distribución de la intuición, y en el capítulo 5, la distribución de las preferencias.
- **Estudio de los cambios de opinión tras la aplicación de las técnicas.** Para aquellas preguntas que se realizaron tanto antes como después del estudio empírico, es necesario comprobar si han ocurrido cambios de opinión. Para ello se van a utilizar tablas de contingencia y el índice de Kappa [4]. Además se va a comprobar si existe direccionalidad en el cambio mediante la Prueba de McNemar-Bowker [5]. Finalmente, en el caso de que el cambio tenga direccionalidad, se comprobará si la distribución de las respuestas en el antes es distinta a la distribución de las respuestas en el después, mediante la Prueba de Stuart-Maxwell [5]. Esto también se realiza en los capítulos 3 y 5 para las intuiciones y preferencias respectivamente.
- **Estudio de la relación entre las distintas preguntas.** Se va a comprobar la relación existente entre las preguntas de intuición entre sí, las preguntas de preferencias entre sí y entre las preguntas de intuición y preferencias. Para esto se va a comprobar si hay acuerdo entre las distintas preguntas, mediante tablas de contingencia y el índice de Kappa. En el capítulo 3, se analiza la relación entre la intuición, en el capítulo 5, la relación entre las preferencias y en el capítulo 8, la relación entre las intuiciones y las preferencias.
- **Estudio de la relación existente entre las intuiciones/preferencias y la realidad.** Finalmente, se comprobará la relación existente entre lo que los sujetos responden y lo que realmente sucede cuando aplican cada una de las técnicas, para esto se va a utilizar el Modelo Lineal General con medidas repetidas [6]. En el capítulo 4, se analiza la relación existente entre las intuiciones y la efectividad real de las técnicas y en el capítulo 6, la relación existente entre las preferencias y la efectividad real de las técnicas.

En la Tabla 2.5 se resumen los test estadísticos realizados y el capítulo en el que se usan.

Capítulo	Distribución	Acuerdo			Realidad
	Prueba de Proporciones (χ^2 de Pearson)	Índice de Acuerdo (Kappa)	Simetría (McNemar-Bowker)	Homogeneidad Marginal (Stuart-Maxwell)	MLG con Medidas Repetidas
3	X	X	X	X	
4					X
5	X	X	X	X	
6					X
7		X			

Tabla 2.5 Test Estadísticos aplicados

A continuación se detallan los test estadísticos utilizados.

2.4.1. Estudio de Intuiciones y Preferencias

Para mayor entendimiento, primero se muestran los resultados en gráficos circulares, donde aparece el porcentaje de sujetos que pertenecen a cada categoría. Para representar el total de sujetos que realizaron el estudio empírico, se han reemplazado los valores nulos por las categorías “Other” y “No answer”, que se explican más adelante.

Por otro lado, también se realizan gráficos circulares eliminado estas categorías, es decir, además de mostrar las proporciones del total de sujetos que realizaron el estudio empírico, también se muestran las proporciones de los sujetos que respondieron únicamente una técnica o programa concreto, dependiendo de si la pregunta está relacionada con las técnicas o con los programas.

Se evalúa cada pregunta (variable respuesta) por separado, por ejemplo, como la pregunta ¿Qué técnica detecta más faltas? (IT1), aparece antes y después, los gráficos circulares se realizan para las dos apariciones y si además en alguna de estas apariciones aparecen respuestas raras, se muestran los gráficos con las categorías (que se hayan creado) y los gráficos sin las categorías.

Para evaluar si la mayoría de los sujetos tienden a responder una técnica en concreto, se ha realizado la prueba de proporciones. Esta prueba se realizó sobre los datos en los que no aparecen las categorías creadas, debido a que las respuestas que interesan son aquellas que tienen una técnica o programa específico, los valores diferentes a estos se han tomado como valores nulos.

Con respecto a las categorías “Other” y “No answer”, debido a que en los cuestionarios las preguntas eran de respuesta abierta, hay varias preguntas a las que los sujetos no responden una técnica o programa concreto. Para estos casos, se ha creado una nueva categoría llamada “Other”. El contenido de esa categoría se explicará en la sección correspondiente, del capítulo 3 o 5, debido a que el contenido y número de sujetos puede variar en cada pregunta e iteración. Finalmente, se indicará cuantos sujetos responden lo mismo y se agrupará el total de sujetos por iteración.

Por otro lado, en ciertas ocasiones los que los sujetos no responden nada, para estos casos, se ha creado una categoría llamada “No answer”. Los sujetos que pertenecen a esta categoría se detallan en cada pregunta, del capítulo 3 o 5, debido a que el número de sujetos puede variar en cada pregunta e iteración. Por ejemplo en la pregunta IT1. Se han creado ambas categorías con los valores mostrados en las Tabla 2.6 y Tabla 2.7.

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: “Depende del caso”	2 sujetos: más de una técnica “E/F”	1 sujeto: más de una técnica “E/F”
	2 sujetos: más de una técnica “E/F”		
	3 sujetos: “Según Programa”		
Total	6 sujetos	2 sujetos	1 sujeto

Tabla 2.6. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	2 sujetos
Iteración 3	3 sujetos

Tabla 2.7. Sujetos en la categoría “No answer”

El test estadístico “Chi-cuadrado de Pearson” permite evaluar si hay diferencia entre las frecuencias observadas en las clases de una variable categórica y las frecuencias esperadas, es decir, indica las diferencias que existen entre una distribución observada y otra teórica.

En nuestro caso, se espera que haya el mismo número de respuestas de cada una de las técnicas/programas. Se espera que el número de sujetos que responden “Revisión”/“Cmdline” sea el mismo que el número de sujetos que responden “Estructural”/“Nametbl” y también el mismo número de sujetos que responden “Funcional”/“Ntree”.

Para empezar es necesario formalizar la hipótesis nula y la hipótesis alternativa:

- H_0 : Las clases de la variable categórica ocurren con igual probabilidad.
($T_1/P_1=T_2/P_2=T_3/P_3$)
- H_1 : Las clases de la variable categórica no ocurren con igual probabilidad.
($T_1/P_1 \neq T_2/P_2 \neq T_3/P_3$)

El nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig} \leq 0.05$) se rechaza la Hipótesis Nula H_0 , en caso contrario ($\text{sig} > 0.05$) no se puede rechazar la Hipótesis Nula H_0 .

2.4.2. Medida de Acuerdo

Una vez estudiadas las proporciones de las preguntas individualmente, es necesario, compararlas dos a dos para poder observar si hay acuerdo entre las mismas, es decir, observar si los sujetos responden lo mismo.

Para el caso de las preguntas que aparecen dos veces interesa saber si las respuestas de antes de realizar el estudio empírico son iguales a las de después de realizarlo. Por otro lado, para el caso en el que se evalúan las preguntas del mismo tipo entre sí y en el que se evalúan las intuiciones con las preferencias, se va a analizar si coincide la respuesta en una pregunta con la respuesta de la otra pregunta.

Antes de proceder a analizar los datos, primero se han organizado los mismos en tablas de contingencia, es decir, tablas en las que se representan las dos preguntas (variables respuesta) a la vez, donde en cada casilla de la tabla aparece el número de casos que pertenecen al mismo tiempo a las dos variables. La tabla de contingencia se define por el número de niveles que puede tomar cada una de las variables. Además para facilitar el entendimiento de los datos, se han representado las tablas de contingencia gráficamente mediante gráficos de barras.

Hay casos en los que los valores totales de la distribución de la técnica no coinciden con los valores totales de las tablas de contingencia, puesto que hay “valores perdidos” para ciertos sujetos, debido a que los sujetos que no responden en alguno de los casos no son los mismos que en el otro caso. Por ejemplo, para mostrar los resultados de la Tabla 2.8 y Tabla 2.9:

- Hay un total de 4 sujetos (los sujetos 10, 18, 26 y 37), en las respuestas antes de aplicar las técnicas, pertenecen a las categorías “Other” y/o “No answer”, sin embargo, después de aplicarlas si hay respuestas.
- Hay un sujeto (el sujeto 11) en las respuestas después de aplicar las técnicas, pertenecen a las categorías “Other” y/o “No answer”, sin embargo antes de aplicarlas si hay respuestas
- Por lo tanto, hay 5 sujetos menos $(46-5)=41$ en las tablas de contingencia, sin embargo, en la distribución de la técnica en el antes hay $(46-4)=42$ sujetos y en el después hay $(46-1)=45$ sujetos.

			After			Total
			Review	Structural	Functional	
Before	Review	Count	5	0	5	10
		% of Total	12,2%	0,0%	12,2%	24,4%
	Structural	Count	2	4	6	12
		% of Total	4,9%	9,8%	14,6%	29,3%
	Functional	Count	6	0	13	19
		% of Total	14,6%	0,0%	31,7%	46,3%
Total		Count	13	4	24	41
		% of Total	31,7%	9,8%	58,5%	100,0%

Tabla 2.8. Ejemplo Tabla de Contingencia

	Antes	Después
Review	10	14
Structural	13	4
Funcional	19	27
Total	42	45

Tabla 2.9. Ejemplo de frecuencias para Antes y Después

Una vez organizados los datos, se va a evaluar si hay acuerdo en las preguntas tal como se explica a continuación.

El acuerdo o desacuerdo (cambio de opinión) se va a medir mediante el índice de concordancia Kappa, visualmente se entiende que hay acuerdo si todos los sujetos se encuentran en la diagonal principal (p_{ij} , con $i=j$) en las tablas de contingencia, sin embargo, dado que este acuerdo puede estar relacionado con el azar, el índice de Kappa permite establecer cuál es el grado de concordancia, teniendo en cuenta el azar, entre las dos respuestas de los sujetos.

Para empezar es necesario formalizar la hipótesis nula y la hipótesis alternativa:

- H_0 : El acuerdo esperado coincide con el azar ($kappa=0$)
- H_1 : El acuerdo esperado no coincide con el azar ($kappa \neq 0$)

El nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($sig \leq 0.05$) se rechaza la Hipótesis Nula H_0 , en caso contrario ($sig > 0.05$) no se puede rechazar la Hipótesis Nula H_0 .

Si se ha rechazado la hipótesis nula, debido a que el valor del índice de Kappa puede tomar valores comprendidos entre -1 y 1, la manera de interpretarlo es la siguiente:

- Si el valor de Kappa es >0 , entonces hay un acuerdo superior al esperado por el azar.
- Si el valor de Kappa es <0 , entonces hay un desacuerdo superior al esperado por el azar.

Tanto en el caso de no poder rechazar la hipótesis nula, como en el que se rechace y el valor de Kappa sea menor que 0, se concluye que no hay acuerdo.

Finalmente para interpretar la relación existente entre el valor de Kappa y el acuerdo se ha propuesto la escala de valoración mostrada en la Tabla 2.10.

Kappa	Grado de Acuerdo/Desacuerdo
=0,00	nulo
>0,00 - 0,20	insignificante
0,21 - 0,40	discreto
>0,41 - 0,60	moderado
0,61 - 0,80	sustancial
0,81 - 1,00	casi perfecto

Tabla 2.10. Grado de acuerdo de Kappa

En el caso de que Kappa sea negativo, se tomará el valor absoluto y se interpretará dicho resultado como “desacuerdo” utilizando la misma tabla.

Finalmente, se puede considerar que el valor de Kappa o acuerdo es *intermedio* cuando su valor es mayor o igual a 0,40 y *bueno* para valores superiores a 0,75[7].

2.4.3. Direccionalidad del desacuerdo

Una vez estudiado el acuerdo, se va a utilizar el test de McNemar-Bowker, este sirve para comprobar la simetría de la tabla de contingencia, es decir, permite comprobar si las casillas situadas simétricamente respecto de la diagonal principal ocurren con la misma probabilidad.

Para empezar es necesario formalizar la hipótesis nula y la hipótesis alternativa:

- H_0 : Los cambios de una categoría a otra se producen en ambas direcciones con igual probabilidad. ($p_{ij}=p_{ji}$)
- H_1 : Los cambios de una categoría a otra no se producen en ambas direcciones con igual probabilidad. ($p_{ij}\neq p_{ji}$)

El nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig} \leq 0.05$) se rechaza la Hipótesis Nula H_0 , en caso contrario ($\text{sig} > 0.05$) no se puede rechazar la Hipótesis Nula H_0 .

En el caso de que haya acuerdo un acuerdo elevado, los valores simétricos a la diagonal serán 0, por lo tanto existirá simetría. Sin embargo aunque no hubiera acuerdo es posible que exista simetría, esto indica que los sujetos cambian de opinión en ambos sentidos, finalmente, si no existe simetría se estudiará la dirección del cambio en cada caso correspondiente.

2.4.4. Homogeneidad Marginal

Una vez estudiada la simetría, se puede contrastar si las diferencias entre las proporciones marginales son significativas. Esto se realiza mediante el test de Stuart-Maxwell que determina si el total de las respuestas de una categoría coinciden en las dos variables analizadas.

Para empezar es necesario formalizar la hipótesis nula y la hipótesis alternativa:

- H_0 : las proporciones marginales son iguales. ($p_{i.}=p_{.i}$)
- H_1 : las proporciones marginales no son iguales. ($p_{i.}\neq p_{.i}$)

El nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig}\leq 0.05$) se rechaza la Hipótesis Nula H_0 , en caso contrario ($\text{sig}>0.05$) no se puede rechazar la Hipótesis Nula H_0 .

En el caso en el que exista acuerdo, las proporciones marginales serán iguales. Por otro lado, aunque no haya acuerdo, en el caso de que exista simetría, también existirá homogeneidad marginal.

En el caso de que exista homogeneidad marginal se espera que la distribución de la técnica de antes sea igual a la de después, sin embargo, es posible, por lo que se comentó en el apartado de acuerdo sobre los “valores perdidos” para ciertos sujetos, que no coincidan el antes con el después.

2.4.5. Relación entre lo respondido y la realidad

Finalmente, es muy importante, evaluar si lo que han respondido los sujetos se corresponde con lo que ha sucedido en realidad, de esta manera se podrá determinar si las intuiciones y las preferencias sirven para tomar decisiones sobre las técnicas de evaluación de código.

Para poder identificar si las variables son independientes, o averiguar cómo afectan a la variable respuesta, se va a utilizar el Análisis de Varianzas (ANOVA). Debido a que todos los sujetos aplicaron todas las técnicas se tiene un factor con medidas repetidas (intra-sujetos), por lo tanto, se va a realizar usando el Modelo Lineal General con Medidas Repetidas.

Para empezar, en nuestro diseño se va a estudiar si el porcentaje de faltas encontradas por todos los sujetos al aplicar cada una de las técnicas (factor intra-sujetos) está relacionado o depende de las respuestas proporcionadas por los sujetos (factor entre-sujetos)

Un factor entre-sujetos, es aquel en el que un grupo de sujetos aparecen en un único nivel del factor y además los sujetos no se repiten en los otros niveles, mientras que en un factor intra-sujetos todos los sujetos aparecen en todos los niveles del factor. Para poder aplicar MLG con medidas repetidas se tienen que cumplir los siguientes requisitos:

- Es necesario comprobar el supuesto de homogeneidad, que asume que las k matrices (una por cada nivel del factor entre-sujetos) son idénticas a nivel poblacional. Esto se va a comprobar con la prueba M de Box, en la cual si el nivel de significancia es mayor que

0.05 ($\text{sig} > 0.05$). no se puede rechazar la hipótesis nula que afirma que las matrices son iguales y por lo tanto se puede aplicar la prueba de Mauchly.

- Otro requisito necesario, es suponer que las varianzas de las diferencias entre cada dos niveles del factor intra-sujetos son iguales. Este supuesto equivale a afirmar que la matriz de varianzas-covarianzas es circular o esférica, por lo tanto es importante comprobar que la matriz de varianzas-covarianzas presente esfericidad. Para ello se va a utilizar la prueba de Mauchly de esfericidad, en la cual, si el nivel de significancia es mayor que 0.05 ($\text{sig} > 0.05$), no se puede rechazar la hipótesis nula que afirma que la matriz de varianzas-covarianzas presenta un patrón de esfericidad. En caso contrario, se evalúan los valores de los test multivariados y en los test univariados, se elegirá uno de los factores de corrección (Épsilon), en nuestro caso se elegirá el estadístico cuyo valor esté más cercano a uno, normalmente suele ser “Huyn-Feldt” que es el menos conservador.

Una vez obtenidos los resultados de las pruebas M de Box y Mauchly se procederá a mirar las tablas del test de efectos intra-sujetos y el test de efectos entre-sujetos.

Para los resultados intra-sujetos, si no se ha aceptado el supuesto de esfericidad se tomarán los valores de los test multivariados y también de los test univariados el estadístico con el valor más cercano a uno, en caso contrario se tomarán los valores de “Esfericidad asumida”. Para cualquiera de los dos casos el nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig} \leq 0.05$) se rechaza la Hipótesis Nula H_0 , que afirma que no hay diferencias entre los distintos niveles del factor. En estos test además de obtener valores para el factor intra-sujetos, también muestran los resultados para la interacción del factor intra-sujetos con el factor entre-sujetos. Se debe comprobar la Hipótesis Nula para los dos casos.

Por otro lado, en los test entre-sujetos, el nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig} \leq 0.05$) se rechaza la Hipótesis Nula H_0 , que afirma que no hay diferencias entre los distintos niveles del factor.

Finalmente, los resultados de MLG se analizarán de la siguiente manera:

- Tanto en el test intra-sujetos como en el entre-sujetos, únicamente en el caso en el que se rechace la Hipótesis Nula, se procederá a analizar la comparación por pares de cada factor, en la cual se muestran las diferencias entre todos los niveles por pares.
 - Para interpretar dichas tablas, primero es necesario comprobar que la diferencia entre las medias sea significativa, es decir, que el nivel de significancia sea menor que 0.05 ($\text{sig} \leq 0.05$), entonces se procede a mirar las diferencias entre las medias. Si esta diferencia es positiva significa que el nivel de la izquierda es mejor que el de la derecha y si es negativa significa que el nivel de la izquierda es peor que el de la derecha.
- Finalmente en el caso en el que el factor entre-sujetos sea significativo, también se procederá a realizar un análisis post-hoc.

- Para el análisis Post-Hoc, primero se realizará el test de Levene para comprobar si hay igualdad de varianzas, el nivel de significancia ha sido establecido al 5% $\rightarrow 0.05$. Por lo tanto, si el nivel de significancia es menor que 0.05 ($\text{sig} \leq 0.05$) se rechaza la Hipótesis Nula H_0 , en el caso en el que no se pueda rechazar la Hipótesis Nula ($\text{sig} > 0.05$) se realizará el test de Tukey, en caso contrario se aplicará el test T2 de Tamhane. Tanto en Tukey como en T2 de Tamhane, los resultados se interpretan igual que en las tablas de comparación por pares.

Para entender mejor los resultados se va a mostrar adicionalmente las medias estimadas y los gráficos asociados a las medias, siempre que sean significativos, cada uno de los dos factores y la interacción de los mismos.

CAPÍTULO 3. RESULTADOS SOBRE INTUICIÓN

En este capítulo se van a presentar y evaluar los resultados obtenidos sobre las intuiciones de los sujetos, tal cual se ha explicado en el capítulo 2. Las preguntas sobre las que se va a trabajar en este capítulo aparecen en la Tabla 3.1.

Código	Preguntas sobre Intuición	It 1	It 2	It 3	Aparición
IT1	¿Qué técnica detecta más faltas?	Si	Si	Si	A/D
IT2	¿Qué técnica has aplicado mejor?	Si	Si	Si	D
IP1	¿Con qué programa has detectado más faltas?	No	Si	Si	D
IP2	¿Con qué programa has detectado menos faltas?	No	No	Si	D

Tabla 3.1. Preguntas sobre Intuición

3.1. Intuición sobre las Técnicas

3.1.1. ¿Qué técnica detecta más faltas?

Dado que esta pregunta se realiza tanto antes como después de la realización del estudio empírico, se va a evaluar si ha habido o no cambios de opinión con respecto a la técnica que los sujetos consideran que detecta más faltas, es decir, cual es la técnica más efectiva. Esta pregunta aparece en las tres iteraciones.

3.1.1.1. Distribución de la Técnica

Primero se va a analizar la distribución de la técnica con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución de la técnica se realizará la prueba de distribución (χ^2 de Pearson).

Se analizará primero tanto antes como después con las categorías “Other” y “No answer” y luego las mismas sin las categorías.

3.1.1.1.1. Resultados de Antes con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 3.2 muestra los resultados de la categoría “Other” y la Tabla 3.3 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: “Depende del caso”	2 sujetos: más de una técnica “E/F”	1 sujeto: más de una técnica “E/F”
	2 sujetos: más de una técnica “E/F”		
	3 sujetos: “Según Programa”		
Total	6 sujetos	2 sujetos	1 sujeto

Tabla 3.2. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	2 sujetos
Iteración 3	3 sujetos

Tabla 3.3. Sujetos en la categoría “No answer”

En la Figura 3.1, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1, se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, sin embargo, hay un gran porcentaje de sujetos que pertenecen a la categoría “Other”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Revisión” y la misma cantidad de sujetos responden “Funcional” y “Estructural”; además hay la misma cantidad de sujetos en la categoría “Other” y “No answer”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Revisión”; además se puede observar que hay más sujetos que no proporcionan ninguna respuesta (categoría “No answer”), que los que responden una cosa distinta (categoría “Other”).



Figura 3.1 Antes con categorías “Other” y “No answer”

3.1.1.1.2. Resultados de Después con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 3.4 muestra los resultados de la categoría “Other” y la Tabla 3.5 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: "Depende del caso"	1 sujeto: más de una técnica "E/F"	
	2 sujetos: más de una técnica "E/F"		
	3 sujetos: "Según Programa"		
Total	6 sujetos	1 sujeto	0 sujetos

Tabla 3.4. Respuestas para categoría "Other"

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 3.5. Sujetos en la categoría "No answer"

En la Figura 3.2, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica "Estructural", sin embargo, hay un gran porcentaje de sujetos que pertenecen a la categoría "Other". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", sin embargo, hay un gran porcentaje de sujetos que pertenecen a la categoría "No answer". Finalmente, para la iteración 3 se puede observar que no existen las categorías "Other" o "No answer" además la mayoría de los sujetos responden la técnica "Revisión".

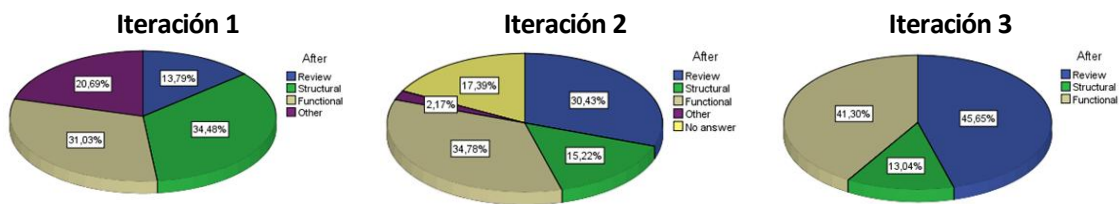


Figura 3.2 Después con categorías "Other" y "No answer"

3.1.1.1.3. Resultados de Antes sin categorías

En esta sección se han puesto a nulo los valores de los sujetos que pertenecían a las categorías: "Other" y "No answer".

En la Figura 3.3, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", seguida de la técnica "Estructural" y finalmente hay muy pocos sujetos que responden la técnica de "Revisión". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Revisión" y la misma cantidad de sujetos responden "Funcional" y "Estructural". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Revisión", seguida de la técnica "Estructural" y finalmente hay muy pocos sujetos que responden la técnica de "Funcional".

Se puede observar que en la iteración 1 la técnica "Funcional" ocurre con mayor probabilidad, mientras que en la iteración 2 y 3 es la técnica "Revisión" la que ocurre con mayor probabilidad.

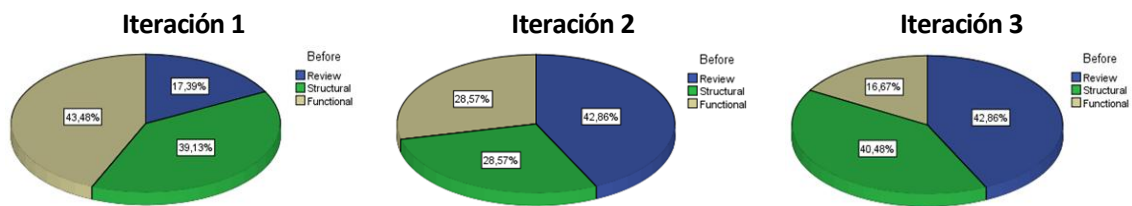


Figura 3.3 Antes sin categorías

3.1.1.1.3.1. Prueba de Proporciones

En la Tabla 3.6, se puede observar que en las tres iteraciones las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que no hay una “intuición definida” o “intuición clara” hacia alguna de las técnicas.

Iteración	Prueba Estadística	gl	Sig. asintótica (bilateral)	N Total
Iteración 1	2,696	2	,260	23
Iteración 2	1,714	2	,424	42
Iteración 3	5,286	2	,071	42

Tabla 3.6. Prueba de Chi-Cuadrado

En la Figura 3.4, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 3.7, para cada una de las iteraciones.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	4	18	18	7,7	14,0	14,0	-3,7	4,0	4,0
Structural	9	12	17	7,7	14,0	14,0	2,3	-2,0	3,0
Functional	10	12	7	7,7	14,0	14,0	1,3	-2,0	-7,0
Total	23	42	42						

Tabla 3.7. Frecuencias para Antes

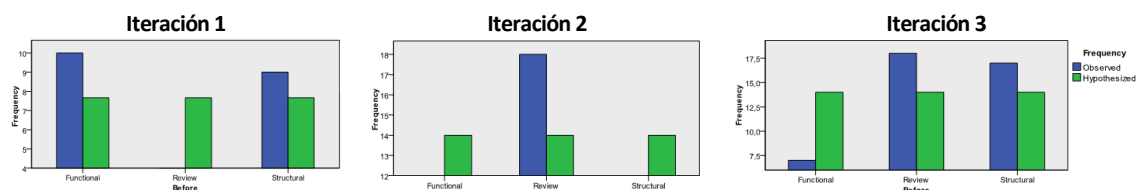


Figura 3.4 Chi-Cuadrado

3.1.1.1.4. Resultados de Después sin Categorías

En esta sección se han puesto a nulo los valores de los sujetos que pertenecían a las categorías: “Other” y “No answer”.

En la Figura 3.5, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica “Estructural”, seguida de la técnica “Funcional” y finalmente hay muy pocos sujetos que responden la técnica de “Revisión”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Finalmente, para la iteración 3 se puede

observar que la mayoría de los sujetos responden la técnica “Revisión”, seguida de la técnica “Funcional” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”.



Figura 3.5 Después sin categorías

Se puede observar que en la iteración 1 la técnica “Estructural” ocurre con mayor probabilidad, mientras que en la iteración 2 es la técnica “Funcional” la que ocurre con mayor probabilidad. Finalmente en la iteración 3 es la técnica “Revisión” la que ocurre con mayor probabilidad.

3.1.1.1.4.1. Prueba de Proporciones

En la Tabla 3.8, se puede observar que para la iteración 1 y 2, las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que no hay una “intuición definida” o “intuición clara” hacia alguna de las técnicas. Mientras que en la iteración 3, las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos intuyen “Revisión” y “Funcional” dado que estas ocurren con mayor probabilidad que la técnica “Estructural”.

Iteración	Prueba Estadística	gl	Sig. asintótica (bilateral)	N Total
Iteración 1	2,696	2	,260	23
Iteración 2	3,622	2	,164	37
Iteración 3	8,652	2	,013	46

Tabla 3.8. Prueba de Chi-Cuadrado

En la Figura 3.6, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 3.9, para cada una de las iteraciones.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	4	14	21	7,7	12,3	15,3	-3,7	1,7	5,7
Structural	10	7	6	7,7	12,3	15,3	1,3	-5,3	-9,3
Functional	9	16	19	7,7	12,3	15,3	2,3	3,7	3,7
Total	23	37	46						

Tabla 3.9. Frecuencias para Después

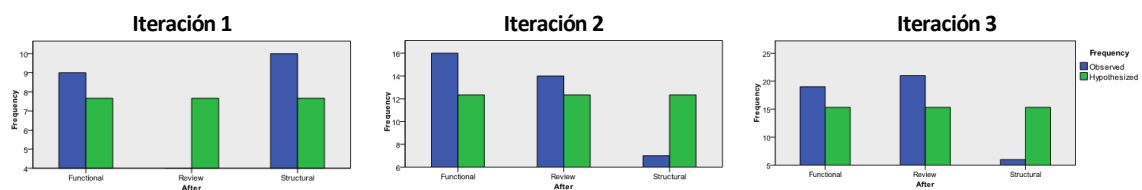


Figura 3.6 Chi-Cuadrado

3.1.1.2. Estudio del Acuerdo Antes / Después

En la Figura 3.7 para la iteración 1, se muestra que de cuatro personas que dicen, antes de la aplicación de las técnicas, que la técnica más efectiva es “Revisión” una cambia de opinión a “Funcional”. De nueve personas que dicen que la técnica más efectiva es “Estructural” cuatro cambian de opinión a “Funcional”. Finalmente, de nueve personas que dicen que la técnica más efectiva es “Funcional” una cambia de opinión a “Revisión” y otras cuatro a “Estructural”.

Para la iteración 2, de doce personas que dicen, antes de la aplicación de las técnicas, que la técnica más efectiva es “Revisión” tres cambian de opinión a “Estructural” y otras tres a “Funcional”. De doce personas que dicen que la técnica más efectiva es “Estructural” cuatro cambian a “Revisión” y otras ocho a “Funcional”. Finalmente se observa que de diez personas que dicen que la técnica más efectiva es “Funcional” tres cambian a “Revisión” y otras tres a “Estructural”.

Para la iteración 3, de dieciocho personas que dicen, antes de la aplicación de las técnicas, que la técnica más efectiva es “Revisión” tres cambian de opinión a “Estructural” y otras siete a “Funcional”. De diecisiete personas que dicen que la técnica más efectiva es “Estructural” ocho cambian a “Revisión” y otras seis a “Funcional”. Finalmente se observa que de siete personas que dicen que la técnica más efectiva es “Funcional” dos cambian de opinión a “Revisión”.

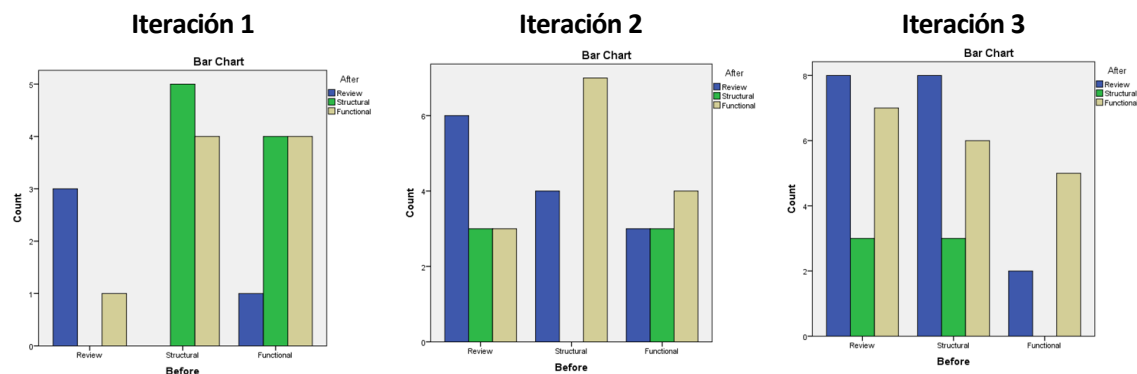


Figura 3.7 Técnica más Efectiva

Estos resultados se pueden contrastar en la Tabla 3.10, Tabla 3.11 y Tabla 3.12 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			After			Total
			Review	Structural	Functional	
Before	Review	Count	3	0	1	4
		% of Total	13,6%	0,0%	4,5%	18,2%
	Structural	Count	0	5	4	9
		% of Total	0,0%	22,7%	18,2%	40,9%
	Functional	Count	1	4	4	9
		% of Total	4,5%	18,2%	18,2%	40,9%
Total		Count	4	9	9	22
		% of Total	18,2%	40,9%	40,9%	100,0%

Tabla 3.10. Tabla de Contingencia: Técnica más Efectiva iteración 1

			After			Total
			Review	Structural	Functional	
Before	Review	Count	6	3	3	12
		% of Total	18,2%	9,1%	9,1%	36,4%
	Structural	Count	4	0	7	11
		% of Total	12,1%	0,0%	21,2%	33,3%
	Functional	Count	3	3	4	10
		% of Total	9,1%	9,1%	12,1%	30,3%
Total		Count	13	6	14	33
		% of Total	39,4%	18,2%	42,4%	100%

Tabla 3.11. Tabla de Contingencia: Técnica más Efectiva iteración 2

			After			Total
			Review	Structural	Functional	
Before	Review	Count	8	3	7	18
		% of Total	19,0%	7,1%	16,7%	42,9%
	Structural	Count	8	3	6	17
		% of Total	19,0%	7,1%	14,3%	40,5%
	Functional	Count	2	0	5	7
		% of Total	4,8%	0,0%	11,9%	16,7%
Total		Count	18	6	18	42
		% of Total	42,9%	14,3%	42,9%	100%

Tabla 3.12. Tabla de Contingencia: Técnica más Efectiva iteración 3

3.1.1.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede analizar el porcentaje de acuerdo sin corrección del azar, el porcentaje de acuerdo para la iteración 1 es del 54.5%, es decir, poco más de la mitad de los sujetos no cambian de opinión. Para la iteración 2 el porcentaje es del 30.3%, muy pocos sujetos mantienen sus respuestas. Finalmente para la iteración 3 el porcentaje es del 38%, muy pocos sujetos mantienen sus respuestas.

En la Tabla 3.13, se puede observar que para las tres iteraciones, el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, existen cambios de opinión. Esto se esperaba, puesto que el porcentaje de acuerdo sin corrección del azar ya es bajo, en las tres iteraciones el porcentaje ronda el 50%.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,281	,180	1,788	,074
Iteración 2	-,044	,109	-,369	,712
Iteración 3	,099	,100	1,041	,282

Tabla 3.13. Índice de Concordancia Kappa

3.1.1.2.2. Direccionalidad en el cambio de opinión

A continuación se muestran los resultados del Test de McNemar-Bowker. Según los resultados mostrados en la Tabla 3.14, en la iteración 1 y en la iteración 2, hay simetría ($\text{sig} > 0.05$) los cambios de una categoría a otra se producen en ambas direcciones con igual probabilidad, es decir, los cambios se producen de forma indistinta. En el caso de la iteración 3, no hay simetría ($\text{sig} \leq 0.05$) los

cambios de una categoría a otra no se producen en ambas direcciones con igual probabilidad, en la Tabla 3.12 se puede observar que los sujetos cambian más de “Estructural” a “Revisión” que de “Revisión” a “Estructural” (8 frente a 3), por otro lado, la mayoría de sujetos cambian hacia “Funcional” independientemente de que previamente hayan contestado “Revisión” (7 frente a 2) o “Estructural” (6 frente a 0).

Iteración	Valor	gl	Sig. asintótica (bilateral)	N de casos válidos
Iteración 1	,000	2	1,000	22
Iteración 2	1,743	3	,627	33
Iteración 3	11,051	3	,011	42

Tabla 3.14. Pruebas de McNemar-Bowker

Resumiendo, los sujetos cambian de “Estructural” hacia “Revisión” (8 sujetos) y “Funcional” (6 sujetos) indistintamente, además hay poco sujetos que cambian de “Revisión” a “Estructural” (3 sujetos) y de “Funcional” a “Revisión” (2 sujetos).

3.1.1.2.3. Homogeneidad Marginal

Finalmente se muestran los resultados del Test de Stuart-Maxwell. Según los resultados mostrados en la Tabla 3.15, para la iteración 1 y la iteración 2, la distribución de las tres técnicas es igual antes que después ($\text{sig} > 0.05$), además se esperaba este resultado puesto que también existe simetría. En el caso de la iteración 3, la distribución de las tres técnicas no es igual antes que después ($\text{sig} \leq 0.05$), este resultado, se esperaba puesto que en la distribución de la técnica, se ha obtenido que en las respuestas del antes las categorías ocurren con la misma probabilidad, sin embargo, las de después no ocurren con la misma probabilidad.

Iteración	χ^2	gl	Sig.
Iteración 1	0	2	1
Iteración 2	1,581	2	0,454
Iteración 3	11,05	2	0,004

Tabla 3.15. Prueba de Stuart-Maxwell

3.1.1.3. Discusión

Finalmente en la Tabla 3.16, se van a evaluar todos los test en conjunto para comprobar si realmente existen cambios de opinión y como son estos cambios.

Para la iteración 1 y la iteración 2, se puede observar que hay cambio de opinión, sin embargo, de este cambio no se obtiene nada claro, puesto que cada sujeto intuye una cosa diferente (tanto antes como después las tres técnicas ocurren la misma probabilidad). Además se puede comprobar que hay simetría, es decir, cuando los sujetos cambian de opinión, el cambio es simétrico ($p_{ij} = p_{ji}$). Finalmente se comprueba que hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica es el mismo.

Para la iteración 3, se puede observar que hay cambio de opinión, con este cambio se obtiene que los sujetos intuyen “Revisión” y “Funcional” (después las tres técnicas no ocurren la misma

probabilidad). Además se puede comprobar que no hay simetría, es decir, cuando los sujetos cambian de opinión, el cambio no es simétrico ($p_{ij} \neq p_{ji}$). Finalmente se comprueba que no hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica, no es el mismo. Esto se puede ver pues se observa que en el caso de las técnicas “Estructural” y “Funcional” el porcentaje cambia considerablemente. Además se puede comprobar con el valor X^2 que tanto para antes como para después las categorías no ocurren con la misma probabilidad.

	Acuerdo		Distribución Técnica				
	Kappa	McNemar-Bowker	Stuart-Maxwell	X^2 (Antes)		X^2 (Después)	
Iteración 1	,074	1,000	1	,260	R: 17,39%	,260	R: 17,39%
					E:39,13%		E:43,48%
					F: 43,48%		F: 39,13%
Iteración 2	,712	,627	0,454	,424	R: 42,86%	,164	R: 37,84%
					E:28,57%		E:18,92%
					F: 28,57%		F: 43,24%
Iteración 3	,282	,011	0,004	,071	R: 42,86%	,013	R: 45,65%
					E:40,48%		E:13,04%
					F: 16,67%		F: 41,30%

Tabla 3.16. Tabla Resumen

3.1.2. ¿Qué técnica has aplicado mejor?

En esta sección se va a evaluar la opinión de los sujetos con respecto a la técnica que consideran que han aplicado mejor. Esta pregunta aparece en las tres iteraciones y solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

3.1.2.1. Distribución de la Técnica

Primero se va a analizar la distribución de la técnica con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución de la técnica se realizará la prueba de distribución (X^2 de Pearson).

3.1.2.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de después en un gráfico circular. La Tabla 3.17 muestra los resultados de la categoría “Other” y la Tabla 3.18 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: más de una técnica “E/F”		
	1 sujeto: “Todas”		
Total	2 sujetos	0 sujetos	0 sujetos

Tabla 3.17. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	1 sujeto
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 3.18. Sujetos en la categoría "No answer"

En la Figura 3.8, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1, se puede observar que la mayoría de los sujetos responden la técnica "Funcional", sin embargo, hay un porcentaje de sujetos en las categorías "Other" y "No answer". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", además se puede observar que hay más sujetos en la categoría "No answer" que en la "Estructural". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Funcional" y se puede observar que no se han creado categorías "Other" y "No answer".



Figura 3.8 Después con categorías

3.1.2.1.2. Resultados sin Categorías

En la Figura 3.9, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", seguida de la técnica "Estructural" y finalmente hay muy pocos sujetos que responden la técnica de "Revisión". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", seguida de la técnica "Revisión" y finalmente hay pocos sujetos que responden la técnica de "Estructural". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", seguida muy de cerca por la técnica "Revisión" y finalmente hay muy pocos sujetos que responden la técnica de "Estructural".

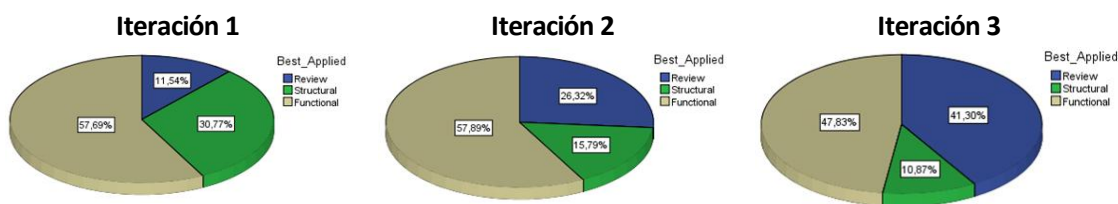


Figura 3.9 Después sin categorías

Se puede observar que en las tres iteraciones la técnica "Funcional" ocurre con mayor probabilidad.

3.1.2.1.2.1. Prueba de Proporciones

En la Tabla 3.19, se puede observar que en las tres iteraciones las categorías no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$).

Para la iteración 1 y 2, los sujetos intuyen “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”, además entre estas dos técnicas no está claro lo que sucede.

Para la iteración 3, los sujetos intuyen “Revisión” y “Funcional” dado que estas ocurren con mayor probabilidad que la técnica “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	26	8,385	2	,015
Iteración 2	38	10,947	2	,004
Iteración 3	46	10,739	2	,005

Tabla 3.19. Prueba de Chi-Cuadrado

En la Figura 3.10, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 3.20, para cada una de las iteraciones.

	N observado			N esperado			Residual		
	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3
Review	3	10	19	8,7	12,7	15,3	-5,7	-2,7	3,7
Structural	8	6	5	8,7	12,7	15,3	-,7	-6,7	-10,3
Functional	15	22	22	8,7	12,7	15,3	6,3	9,3	6,7
Total	26	38	46						

Tabla 3.20. Frecuencias

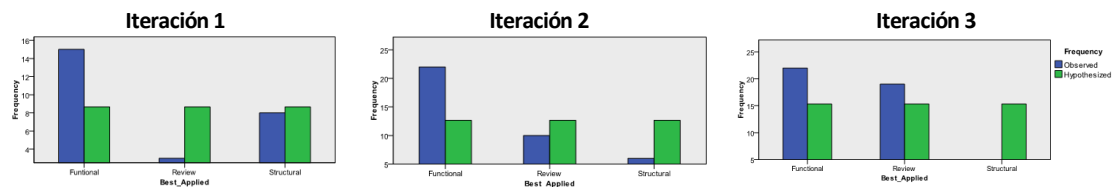


Figura 3.10 Chi-Cuadrado

3.2. Intuición sobre los Programas

3.2.1. ¿Con qué programa has detectado más faltas?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa en el que consideran que han detectado más faltas. Esta pregunta no aparece en la primera iteración, únicamente aparece en la segunda y tercera iteración. Solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

3.2.1.1. Distribución del Programa

Primero se va a analizar la distribución de los programas con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución del programa se realizará la prueba de distribución (χ^2 de Pearson).

3.2.1.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de después en un gráfico circular. La Tabla 3.21 muestra los resultados de la categoría “Other” y la Tabla 3.22 muestra los resultados de la categoría “No answer”

	Iteración 2	Iteración 3
Respuestas		1 sujeto: más de un programa “Nametbl/Ntree”
Total	0 sujetos	1 sujeto

Tabla 3.21. Respuestas para categoría "Other"

Iteración	Nº de sujetos
Iteración 2	9 sujetos
Iteración 3	0 sujetos

Tabla 3.22. Sujetos en la categoría "No answer"

En la Figura 3.11, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa “Cmdline”, además se puede observar que hay un gran porcentaje de sujetos en la categoría “No answer”. Finalmente, para la iteración 3 se puede observar que al igual que en la iteración 2 la mayoría de los sujetos responden el programa “Cmdline”, y hay muy pocos sujetos que no responden, es decir, pertenecen a la categoría “Other”.

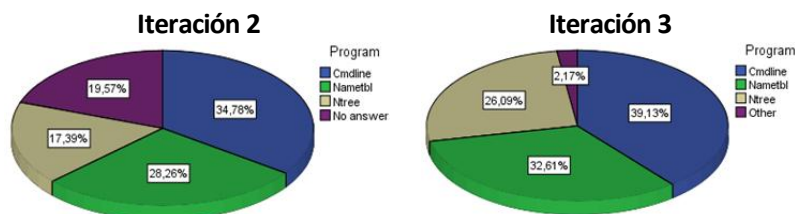


Figura 3.11 Después con categorías

3.2.1.1.2. Contraste de proporciones sin categorías

En la Figura 3.12, para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa “Cmdline”, seguido de “Nametbl” y finalmente hay pocos sujetos que responden el programa “Ntree”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa “Cmdline”, seguido de “Nametbl” y finalmente hay pocos sujetos que responden el programa “Ntree”. En ambas iteraciones el programa “Cmdline” ocurre con mayor probabilidad.

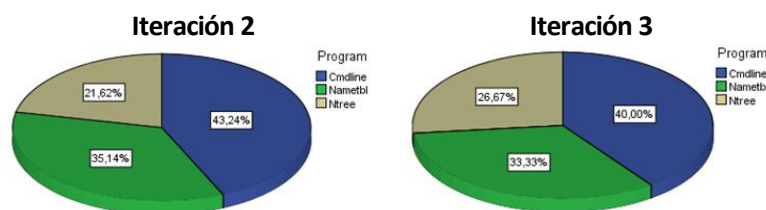


Figura 3.12 Después sin categorías

3.2.1.1.2.1. Prueba de proporciones

En la Tabla 3.23. Prueba de Chi-Cuadrado, se puede observar que en las dos iteraciones, los tres programas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que no hay una “intuición definida” o “intuición clara” hacia alguno de los programas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 2	37	2,649	2	,266
Iteración 3	45	1,200	2	,549

Tabla 3.23. Prueba de Chi-Cuadrado

En la Figura 3.13, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de los programas mostradas en la Tabla 3.24, para cada una de las iteraciones.

	N observado		N esperado		Residual	
	It. 2	It. 3	It. 2	It. 3	It. 2	It. 3
Cmdline	16	18	12,3	15,0	3,7	3,0
Nametbl	13	15	12,3	15,0	,7	,0
Ntree	8	12	12,3	15,0	-4,3	-3,0
Total	37	45				

Tabla 3.24. Frecuencias

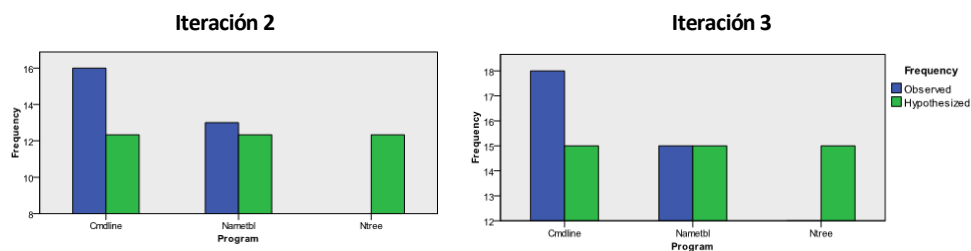


Figura 3.13 Chi-Cuadrado

3.2.2. ¿Con qué programa has detectado menos faltas?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa en el que consideran que han detectado menos faltas. Esta pregunta no aparece ni en la primera, ni en la segunda iteración, únicamente aparece en la tercera iteración. Solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

3.2.2.1. Distribución del Programa

Debido a que no existen respuestas en las que no hay un único programa, ni tampoco respuestas en blanco, no es necesario crear categorías.

En la Figura 3.14, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa “Ntree”, y la misma cantidad de sujetos responden “Cmdline” y “Nametbl”.

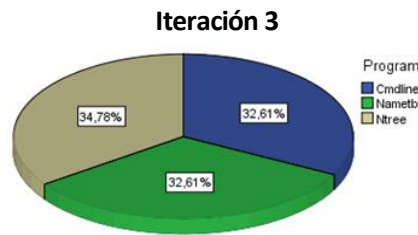


Figura 3.14 Después sin categorías

3.2.2.1.1. Prueba de proporciones

En la Tabla 3.25, se puede observar que los tres programas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que no hay una “intuición definida” o “intuición clara” hacia alguno de los programas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 3	46	,043	2	,978

Tabla 3.25. Prueba de Chi-Cuadrado

En la Figura 3.15, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de los programas mostrados en la Tabla 3.26, para la iteración 3.

	N observado	N esperado	Residual
Cmdline	15	15,3	-,3
Nametbl	15	15,3	-,3
Ntree	16	15,3	,7
Total	46		

Tabla 3.26. Frecuencias

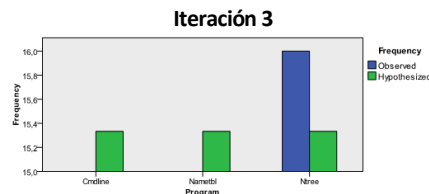


Figura 3.15 Chi-Cuadrado

3.3. Consistencia en las Contestaciones sobre las Técnicas

En este capítulo se va a analizar la relación existente entre las preguntas de intuición entre sí, para ello se seguirá el método explicado en el capítulo 2. Se va a comparar cada variable de la fila de la izquierda con las variables que aparecen en la derecha, en la Tabla 3.27.

Pregunta 1	Pregunta 2
Técnica más efectiva	Técnica mejor aplicada
Programa más efectivo	Programa menos efectivo

Tabla 3.27. Relación entre las preguntas de intuición

3.3.1. ¿Qué técnica detecta más faltas? / ¿Qué técnica has aplicado mejor?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que detecte más faltas y la técnica que han aplicado mejor. Esta pregunta aparece en las tres iteraciones. Dado que la segunda pregunta solo se realizó después del estudio empírico, solo se va a analizar la combinación: Después-Después.

3.3.1.1. Estudio del Acuerdo Después-Después

En la Figura 3.16, para la iteración 1, se observa que de tres personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, ninguna la considera la técnica mejor aplicada, por lo tanto los tres sujetos se distribuyen así: para dos la técnica mejor aplicada es “Estructural” y para uno “Funcional”. De nueve personas que dicen que la técnica más efectiva es “Estructural” tres también la consideran la técnica mejor aplicada, de los seis restantes: uno considera que la mejor aplicada es “Revisión” y los otros cinco “Funcional”. Finalmente, de nueve personas que dicen que la técnica más efectiva es “Funcional” siete también la consideran la técnica mejor aplicada y los otros dos restantes consideran que la mejor aplicada es “Revisión”.

Para la iteración 2, de catorce personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, ocho también la consideran la técnica mejor aplicada, de los seis restantes: para dos la mejor aplicada es “Estructural” y para cuatro es “Funcional”. De siete personas que dicen que la técnica más efectiva es “Estructural” tres también la consideran la técnica mejor aplicada, de los cuatro restantes: dos consideran que la mejor aplicada es “Revisión” y los otros dos es “Funcional”. Finalmente se observa que de dieciséis personas que dicen que la técnica más efectiva es “Funcional” los dieciséis también la consideran la técnica mejor aplicada.

Para la iteración 3, de veintiuno personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, quince también la consideran la técnica mejor aplicada y para seis la técnica mejor aplicada es “Funcional”. De seis personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran la técnica mejor aplicada y para cuatro la técnica mejor aplicada es “Funcional”. Finalmente se observa que de diecinueve personas que dicen que la técnica más efectiva es “Funcional” doce también la consideran la técnica mejor aplicada, de los siete restantes: para cuatro la técnica mejor aplicada es “Revisión” y para tres “Estructural”.

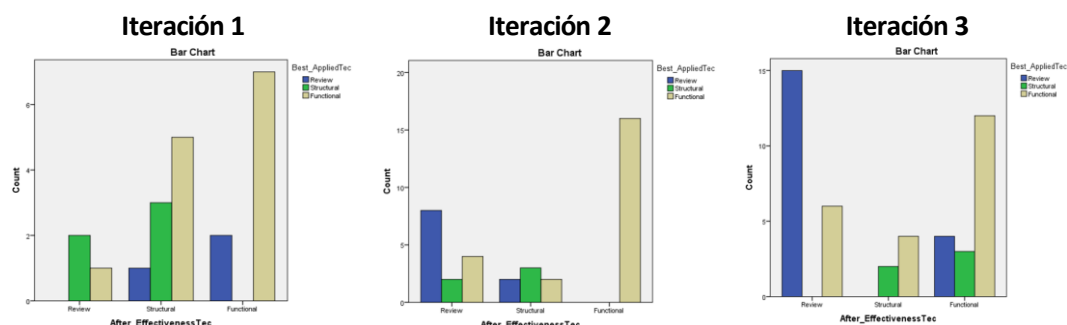


Figura 3.16 Técnica más Efectiva / Técnica mejor Aplicada

Estos resultados se pueden contrastar en la Tabla 3.28, Tabla 3.29 y Tabla 3.30 para la iteración 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			IT2			Total
			Review	Structural	Functional	
IT1 (After)	Review	Count	0	2	1	3
		% of Total	0,0%	9,5%	4,8%	14,3%
	Structural	Count	1	3	5	9
		% of Total	4,8%	14,3%	23,8%	42,9%
	Functional	Count	2	0	7	9
		% of Total	9,5%	0,0%	33,3%	42,9%
Total		Count	3	5	13	21
		% of Total	14,3%	23,8%	61,9%	100,0%

Tabla 3.28. Tabla de Contingencia de la iteración 1

			IT2			Total
			Review	Structural	Functional	
IT1 (After)	Review	Count	8	2	4	14
		% of Total	21,6%	5,4%	10,8%	37,8%
	Structural	Count	2	3	2	7
		% of Total	5,4%	8,1%	5,4%	18,9%
	Functional	Count	0	0	16	16
		% of Total	0,0%	0,0%	43,2%	43,2%
Total		Count	10	5	22	37
		% of Total	27,0%	13,5%	59,5%	100,0%

Tabla 3.29. Tabla de Contingencia de la iteración 2

			IT2			Total
			Review	Structural	Functional	
IT1 (After)	Review	Count	15	0	6	21
		% of Total	32,6%	0,0%	13,0%	45,7%
	Structural	Count	0	2	4	6
		% of Total	0,0%	4,3%	8,7%	13,0%
	Functional	Count	4	3	12	19
		% of Total	8,7%	6,5%	26,1%	41,3%
Total		Count	19	5	22	46
		% of Total	41,3%	10,9%	47,8%	100,0%

Tabla 3.30. Tabla de Contingencia de la iteración 3

3.3.1.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 47.6%, es decir menos de la mitad de los sujetos están de acuerdo en las dos preguntas. Para la iteración 2 el porcentaje es del 72.9%, es decir la mayoría de los sujetos están de acuerdo en las dos preguntas. Finalmente para la iteración 3 el porcentaje es del 63%, es decir la mayoría de los sujetos están de acuerdo en las dos preguntas.

En la Tabla 3.31, para la iteración 1 se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo. Para la iteración 2 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa muestra que el acuerdo es moderado ($> 0,41 - 0,60$), se puede concluir que existe acuerdo. Finalmente, para la iteración 3 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa muestra que el acuerdo es discreto ($0,21 - 0,40$), se puede concluir que existe acuerdo. Esto se esperaba, puesto que el porcentaje de acuerdo sin corrección del azar para la iteración 1 ya es bajo, ronda el 50%. Mientras que en la iteraciones 2 y 3, el porcentaje es superior al 50%.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,144	,139	,939	,348
Iteración 2	,561	,109	4,756	,000
Iteración 3	,384	,114	3,350	,001

Tabla 3.31. Índice de Concordancia Kappa

Debido a que en la iteración 1, no hay acuerdo, en la iteración 2, el acuerdo es moderado y en la iteración 3 el acuerdo es discreto, se puede concluir que los sujetos identifican la técnica que detecta más faltas con la técnica mejor aplicada, sin embargo, la relación entre las preguntas es intermedia.

3.4. Consistencia en las Contestaciones sobre los Programas

3.4.1. ¿Con qué programa has detectado más faltas? / ¿Con qué programa has detectado menos faltas?

En esta sección se va a evaluar la relación existente entre el programa que los sujetos consideran que es en el que han detectado más faltas y en el que consideran que han detectado menos faltas. Esta pregunta solo aparece en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después. Debido a que las preguntas son contradictorias, se espera que no haya acuerdo entre las mismas.

3.4.1.1. Estudio del Acuerdo

En la Figura 3.17, para la iteración 3, de dieciocho personas que dicen que el programa en el que detectaron más faltas es "Cmdline", uno también lo considera el programa menos efectivo, de los diecisiete restantes: nueve consideran que el programa menos efectivo es "Nametbl" y ocho "Ntree". De quince personas que dicen que el programa en el que detectaron más faltas es "Nametbl", uno también lo considera el programa menos efectivo, de los catorce restantes: seis consideran que el menos efectivo es "Cmdline" y ocho "Ntree". Finalmente doce personas que dicen que el programa en el que detectaron más faltas es "Ntree", ninguna lo considera el programa menos efectivo, por lo tanto de los doce sujetos: siete consideran que el menos efectivo es "Cmdline" y cinco "Nametbl".

Estos resultados se pueden contrastar en la Tabla 3.32. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

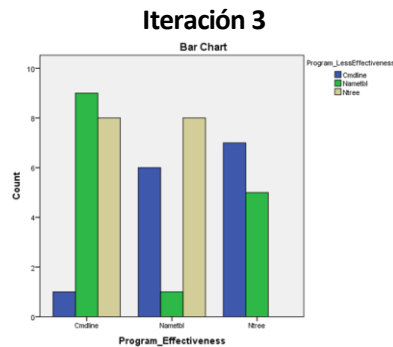


Figura 3.17 Programa más Efectivo / Programa menos Efectivo

			IP2			Total
			Cmdblne	Nametbl	Ntree	
IP1	Cmdblne	Count	11	4	3	18
		% of Total	27,5%	10,0%	7,5%	45,0%
	Nametbl	Count	5	2	8	15
		% of Total	12,5%	5,0%	20,0%	37,5%
	Ntree	Count	3	2	2	7
		% of Total	7,5%	5,0%	5,0%	17,5%
Total		Count	19	8	13	40
		% of Total	47,5%	20,0%	32,5%	100,0%

Tabla 3.32. Tabla de Contingencia de la iteración 3

3.4.1.1.1. Medida de Acuerdo

En la tabla de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 37.8%, es decir muy pocos sujetos no cambian de opinión.

En la Tabla 3.33, se observa que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa es negativo se interpretará el valor como desacuerdo, es decir, existe un desacuerdo moderado ($>0,41 - 0,60$), se puede concluir que no existe acuerdo entre las dos preguntas. Esto se esperaba, puesto que el porcentaje de acuerdo sin corrección del ya es bajo

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	-,427	,048	-4,108	,000

Tabla 3.33. Índice de Concordancia Kappa

Se puede concluir dado que el desacuerdo es moderado, no hay acuerdo, por lo tanto los sujetos no identifican el programa más con el programa en el menos efectivo.

CAPÍTULO 4. COMPARANDO LAS INTUICIONES CON LA REALIDAD

El objetivo de este capítulo es responder a la pregunta: ¿sirve la intuición para tomar decisiones sobre testing? Concretamente sobre la efectividad de las técnicas de testing.

Para ello, se estudia si ha habido concordancias para cada sujeto entre lo que intuían y los resultados cuantitativos del estudio empírico. Esto se hace comparando los resultados sobre intuición con los resultados cuantitativos, tal cual muestra la Tabla 4.1.

Resultados sobre Intuición	Resultados Cuantitativos
Técnica más efectiva	% defectos encontrados
Técnica mejor aplicada	
Programa más efectivo	
Programa menos efectivo	

Tabla 4.1 Comparación Intuición con Realidad

A partir de los resultados se podrá concluir si la intuición de los sujetos sirve para predecir la efectividad de las técnicas.

4.1. Intuiciones sobre las Técnicas

4.1.1. ¿Qué técnica detecta más faltas?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y la técnica que los sujetos consideran que es la que detecta más faltas. Esta pregunta aparece en las tres iteraciones y se van a evaluar las respuestas que dieron los sujetos tanto antes como después de aplicar las técnicas.

4.1.1.1. Respuestas Antes

4.1.1.1.1. Iteración 1

En la Tabla 4.2, se puede observar que para el caso de las técnicas “Funcional” y “Estructural” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	64,2850	18,44056	4
	Structural	57,1400	40,40408	4
	Functional	67,8550	24,70480	4
Structural	Review	52,3822	27,66257	9
	Structural	83,8611	11,15270	9
	Functional	79,8933	15,20326	9
Functional	Review	51,4280	26,25390	10
	Structural	54,0470	40,09293	10
	Functional	89,7590	7,10503	10

Tabla 4.2. Estadísticos Descriptivos

En la Tabla 4.3, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	23,423
F	1,345
df1	12
df2	404,023
Sig.	,190

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

Tabla 4.3. Test de igualdad de matrices de covarianzas

En la Tabla 4.4, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,764	5,120	2	,077	,809	,958	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.4. Test de esfericidad de Mauchly

El siguiente paso es analizar los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 4.5, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	5311,290	2	2655,645	3,681	,034
AppliedTec * IT1(Before)	Sphericity Assumed	5600,597	4	1400,149	1,941	,122
Error(AppliedTec)	Sphericity Assumed	28859,206	40	721,480		

Tabla 4.5. Test univariado de efectos intra-sujetos

Finalmente, en la Tabla 4.6, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	260805,704	1	260805,704	580,856	,000
IT1(Before)	966,769	2	483,384	1,077	,360
Error	8980,046	20	449,002		

Tabla 4.6. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- Para los casos en los que se tiene que aplicar conocimiento recién adquirido, la intuición no sirve para predecir. Por lo tanto, no garantiza que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 4.7, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y la más baja los que aplicaron “Revisión”. Esto se puede observar gráficamente en la Figura 4.1.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	56,032	5,847	43,835	68,228
Structural	65,016	7,222	49,951	80,081
Functional	79,169	3,254	72,381	85,958

Tabla 4.7. Medias estimadas

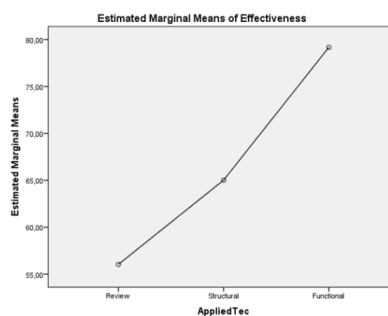


Figura 4.1 Técnica Aplicada

Finalmente, en la Tabla 4.8, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Además, no existen diferencias significativas entre la

“Revisión” y la técnica “Estructural” y por último, tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	-8,984	10,444	1,000	-36,271	18,303
	Functional	-23,137*	7,066	,011	-41,598	-4,677
Structural	Review	8,984	10,444	1,000	-18,303	36,271
	Functional	-14,153	7,923	,268	-34,853	6,547
Functional	Review	23,137*	7,066	,011	4,677	41,598
	Structural	14,153	7,923	,268	-6,547	34,853

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.8. Comparación por pares

4.1.1.1.2. Iteración 2

En la Tabla 4.9, se puede observar que para el caso de las técnicas “Revisión” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	61,1111	21,38963	18
	Structural	45,3704	15,97133	18
	Functional	42,5924	19,15053	18
Structural	Review	56,9444	16,60342	12
	Structural	50,0000	20,10076	12
	Functional	45,8333	18,96967	12
Functional	Review	45,8333	24,74619	12
	Structural	48,6111	16,60342	12
	Functional	54,1667	14,43376	12

Tabla 4.9. Estadísticos Descriptivos

En la Tabla 4.10, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	9,498
F	,698
df1	12
df2	5496,456
Sig.	,755

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

Tabla 4.10. Test de igualdad de matrices de covarianzas

En la Tabla 4.11, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,991	,332	2	,847	,991	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.11. Test de esfericidad de Mauchly

El siguiente paso es analizar los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 4.12, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	1277,662	2	638,831	1,906	,156
AppliedTec * IT1(Before)	Sphericity Assumed	2821,161	4	705,290	2,104	,088
Error(AppliedTec)	Sphericity Assumed	26147,220	78	335,221		

Tabla 4.12. Test univariado de efectos intra-sujetos

Finalmente, en la Tabla 4.13, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	304375,071	1	304375,071	756,237	,000
IT1(Before)	43,726	2	21,863	,054	,947
Error	15696,977	39	402,487		

Tabla 4.13. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- Para los casos en los que se tiene que aplicar conocimiento recién adquirido, la intuición no sirve para predecir. Por lo tanto, no garantiza que una técnica se vaya a aplicar mejor que otra.

4.1.1.1.3. Iteración 3

En la Tabla 4.14, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” no coincide ya que la que mejor aplican es “Revisión” y en el caso de “Funcional” tampoco coincide ya que la que mejor aplican es “Estructural”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	55,5611	19,80957	18
	Structural	43,5222	19,91359	18
	Functional	41,6667	14,28928	18
Structural	Review	57,8471	21,34520	17
	Structural	37,2529	25,35437	17
	Functional	42,1471	11,96778	17
Functional	Review	40,4714	23,30556	7
	Structural	54,7429	20,88898	7
	Functional	40,4714	16,27316	7

Tabla 4.14. Estadísticos Descriptivos

En la Tabla 4.15, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	18,636
F	1,321
df1	12
df2	1638,401
Sig.	,199

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

Tabla 4.15. Test de igualdad de matrices de covarianzas

En la Tabla 4.16, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,891	4,396	2	,111	,902	,991	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.16. Test de esfericidad de Mauchly

El siguiente paso es analizar los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 4.17, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	1735,313	2	867,656	2,581	,082
AppliedTec * IT1(Before)	Sphericity Assumed	3063,533	4	765,883	2,278	,068
Error(AppliedTec)	Sphericity Assumed	26221,441	78	336,172		

Tabla 4.17. Test univariado de efectos intra-sujetos

Finalmente, en la Tabla 4.18, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	221759,118	1	221759,118	480,609	,000
Before_ IT1(Before)	57,749	2	28,875	,063	,939
Error	17995,085	39	461,412		

Tabla 4.18. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- Para los casos en los que se tiene que aplicar conocimiento recién adquirido, la intuición no sirve para predecir. Por lo tanto, no garantiza que una técnica se vaya a aplicar mejor que otra.

4.1.1.1.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 4.19, se observa que únicamente en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas que en “Revisión”.

En el resto de los casos no influye ni la técnica aplicada, ni la intuición que se tiene sobre la técnica, ni la interacción de las dos.

Debido a que únicamente en una iteración hay técnicas que detectan más faltas, se podría decir, que realmente no es la técnica la que determina este comportamiento. Es decir, al parecer existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que influyen al aplicar las técnicas. Finalmente, está claro que la intuición no sirve para predecir la efectividad de las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,034	F>R	,360	N/A	,122	N/A
		R=E				
		E=F				
Iteración 2	,156	N/A	,947	N/A	,088	N/A
Iteración 3	,082	N/A	,939	N/A	,068	N/A

Tabla 4.19. Resumen

4.1.1.2. Respuestas Después

4.1.1.2.1. Iteración 1

En la Tabla 4.20, se puede observar que para el caso de la técnica “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” y “Estructural” no coincide ya que en ambos casos la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (After)	AppliedTec	Mean	Std. Desviation	N
Review	Review	64,2850	18,44056	4
	Structural	57,1400	40,40408	4
	Functional	67,2600	24,15395	4
Structural	Review	44,2860	27,31149	10
	Structural	70,9510	29,11169	10
	Functional	85,2360	14,09938	10
Functional	Review	57,1433	23,68899	9
	Structural	65,0789	38,53860	9
	Functional	85,1833	11,02470	9

Tabla 4.20. Estadísticos Descriptivos

En la Tabla 4.21, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	8,182
F	,470
df1	12
df2	404,023
Sig.	,932

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

Tabla 4.21. Test de igualdad de matrices de covarianzas

En la Tabla 4.22, se observa que también es posible asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,768	5,020	2	,081	,812	,961	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.22. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 4.23, se observa que el porcentaje de faltas encontradas no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	5720,862	2	2860,431	3,572	,037
AppliedTec * IT1(After)	Sphericity Assumed	2726,306	4	681,576	,851	,501
Error(AppliedTec)	Sphericity Assumed	32031,047	40	800,776		

Tabla 4.23. Test univariado de efectos intra-sujetos

Finalmente, en la Tabla 4.24, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig}>0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	257268,489	1	257268,489	564,797	,000
IT1(After)	326,090	2	163,045	,358	,704
Error	9110,117	20	455,506		

Tabla 4.24. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- La intuición no sirve para deducir lo que ha sucedido en realidad. Por lo tanto, no garantiza que una técnica se haya aplicado mejor que otra.

Para la técnica aplicada, en la Tabla 4.25, se observa que la media más alta y más la tienen los sujetos que aplicaron “Funcional” y “Revisión”, respectivamente. Esto se observa en la Figura 4.2.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	55,238	5,596	43,566	66,910
Structural	64,390	7,907	47,896	80,884
Functional	79,226	3,400	72,135	86,318

Tabla 4.25. Medias estimadas

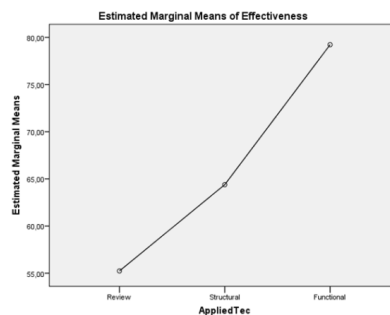


Figura 4.2. Técnica Aplicada

Además, en la Tabla 4.26, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias entre “Revisión” y “Estructural”. Finalmente tampoco existen diferencias entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	-9,152	10,799	1,000	-37,366	19,062
	Functional	-23,988*	6,939	,007	-42,117	-5,859
Structural	Review	9,152	10,799	1,000	-19,062	37,366
	Functional	-14,836	9,022	,347	-38,406	8,733
Functional	Review	23,988*	6,939	,007	5,859	42,117
	Structural	14,836	9,022	,347	-8,733	38,406

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.26. Comparación por pares

4.1.1.2.2. Iteración 2

En la Tabla 4.27, se puede observar que para el caso de la técnica “Revisión” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	63,0952	14,87637	14
	Structural	46,4286	19,80522	14
	Functional	48,8095	19,01898	14
Structural	Review	54,7619	28,40560	7
	Structural	42,8571	13,11326	7
	Functional	42,8567	21,20722	7
Functional	Review	42,7083	24,31944	16
	Structural	47,9167	18,13019	16
	Functional	48,9583	14,23188	16

Tabla 4.27. Estadísticos Descriptivos

En la Tabla 4.28, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	23,706
F	1,662
df1	12
df2	1767,638
Sig.	,069

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

Tabla 4.28. Test de igualdad de matrices de covarianzas

En la Tabla 4.29, se puede observar que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,989	,354	2	,838	,989	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.29. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos Con respecto a la técnica aplicada, en la Tabla 4.30, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	1150,333	2	575,166	1,551	,220
AppliedTec * IT1(After)	Sphericity Assumed	2479,507	4	619,877	1,671	,167
Error(AppliedTec)	Sphericity Assumed	25223,336	68	370,931		

Tabla 4.30. Test univariado de efectos intra-sujetos

Finalmente, en la Tabla 4.31, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	231452,070	1	231452,070	611,593	,000
IT1(After)	991,810	2	495,905	1,310	,283
Error	12867,005	34	378,441		

Tabla 4.31. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- La intuición no sirve para deducir lo que ha sucedido en realidad. Por lo tanto, no garantiza que una técnica se haya aplicado mejor que otra.

4.1.1.2.3. Iteración 3

En la Tabla 4.32, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

IT1 (After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	60,3333	17,06184	21
	Structural	36,5048	22,73998	21
	Functional	38,0857	14,08536	21
Structural	Review	55,5333	22,76775	6
	Structural	44,4667	22,76775	6
	Functional	36,1167	16,37186	6
Functional	Review	45,6105	22,12414	19
	Structural	45,6053	25,35154	19
	Functional	43,8526	13,86175	19

Tabla 4.32. Estadísticos Descriptivos

En la Tabla 4.33, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$). En la Tabla 4.34, se puede observar que no es posible asumir esfericidad ($\text{sig} \leq 0.05$).

El siguiente paso es analizar los test multivariados y en los test univariados valor de Huyn-Feldt. Con respecto a la técnica aplicada, tantee la Tabla 4.35 como en la Tabla 4.36, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, en la Tabla 4.35, se puede observar que la influencia es significativa

(sig<0.05). Sin embargo, en la Tabla 4.36, se puede observar que la influencia no es significativa (sig>0.05).

Box's M ^a	15,704
F	1,093
df1	12
df2	988,994
Sig.	,362

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

Tabla 4.33. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,817	8,476	2	,014	,845	,917	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.34. Test de esfericidad de Mauchly

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,268	7,702 ^b	2,000	42,000	,001
	Wilks' Lambda	,732	7,702 ^b	2,000	42,000	,001
	Hotelling's Trace	,367	7,702 ^b	2,000	42,000	,001
	Roy's Largest Root	,367	7,702 ^b	2,000	42,000	,001
AppliedTec * IT1(After)	Pillai's Trace	,210	2,519	4,000	86,000	,047
	Wilks' Lambda	,793	2,583 ^b	4,000	84,000	,043
	Hotelling's Trace	,258	2,641	4,000	82,000	,039
	Roy's Largest Root	,244	5,235 ^c	2,000	43,000	,009

a. Design: Intercept + IT1(After)

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 4.35. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3966,459	1,691	2345,647	5,601	,008
	Huynh-Feldt	3966,459	1,835	2162,032	5,601	,007
	Lower-bound	3966,459	1,000	3966,459	5,601	,023
AppliedTec * IT1(After)	Greenhouse-Geisser	3527,644	3,382	1043,072	2,491	,060
	Huynh-Feldt	3527,644	3,669	961,422	2,491	,055
	Lower-bound	3527,644	2,000	1763,822	2,491	,095
Error(AppliedTec)	Greenhouse-Geisser	30452,882	72,712	418,812		
	Huynh-Feldt	30452,882	78,888	386,028		
	Lower-bound	30452,882	43,000	708,207		

Tabla 4.36. Test univariado de efectos intra-sujetos

Todos los estadísticos de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$). Pero para la interacción de la técnica aplicada con la elegida no coinciden los test multivariados ($\text{sig} \leq 0.05$) con los test univariados ($\text{sig} > 0.05$).

En la Tabla 4.37, se observa que las opiniones de la técnica más efectiva no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	205961,972	1	205961,972	441,910	,000
IT1(After)	2,267	2	1,134	,002	,998
Error	20041,089	43	466,072		

Tabla 4.37. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- La intuición sirve para deducir lo que ha sucedido en realidad. Por lo tanto, puede garantizar que una técnica se haya aplicado mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 4.38, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y la media más baja “Funcional”. Esto se puede observar gráficamente en la Figura 4.3.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	53,826	3,447	46,875	60,777
Structural	42,192	4,111	33,902	50,483
Functional	39,352	2,459	34,393	44,311

Tabla 4.38. Medias estimadas

Finalmente, en la Tabla 4.39, se puede comprobar que la técnica “Revisión” es significativamente mejor que la “Funcional”. Además, no existen diferencias significativas entre la “Revisión” y la técnica “Estructural” y por último, tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	11,634	5,370	,108	-1,744	25,011
	Functional	14,474*	3,645	,001	5,394	23,554
Structural	Review	-11,634	5,370	,108	-25,011	1,744
	Functional	2,841	4,571	1,000	-8,546	14,228
Functional	Review	-14,474*	3,645	,001	-23,554	-5,394
	Structural	-2,841	4,571	1,000	-14,228	8,546

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.39. Comparación por pares

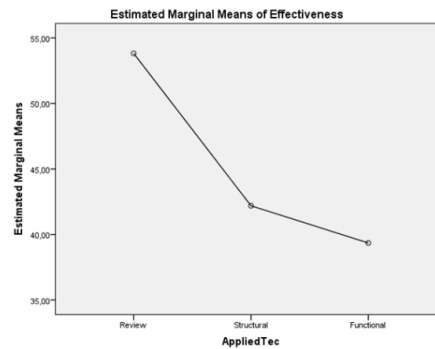


Figura 4.3. Técnica Aplicada

Con respecto a la interacción técnica elegida con técnica aplicada, en la Tabla 4.40, se puede observar que cuando eligen “Revisión” coincide con la técnica que aplicaron mejor. Sin embargo, cuando eligen “Estructural” y “Funcional” no coinciden con la que aplicaron mejor, puesto que la técnica que aplicaron mejor es “Revisión”.

IT1(After)	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	60,333	4,367	51,525	69,141
	Structural	36,505	5,209	26,000	47,010
	Functional	38,086	3,116	31,802	44,369
Structural	Review	55,533	8,171	39,055	72,011
	Structural	44,467	9,745	24,813	64,120
	Functional	36,117	5,829	24,361	47,872
Functional	Review	45,611	4,592	36,351	54,870
	Structural	45,605	5,476	34,561	56,649
	Functional	43,853	3,276	37,246	50,459

Tabla 4.40. Medias estimadas

En la Figura 4.4, se puede observar que los que respondieron “Revisión” aplican bastante peor las técnicas “Estructural” y “Funcional” y a estas dos técnicas las aplican más o menos igual. En el caso de los que responden “Estructural” la que peor aplican es “Funcional” y la que mejor aplican en “Revisión”. Finalmente los que eligen “Funcional” se observa que aplican mejor “Revisión”, aunque no hay grandes diferencias entre las tres.

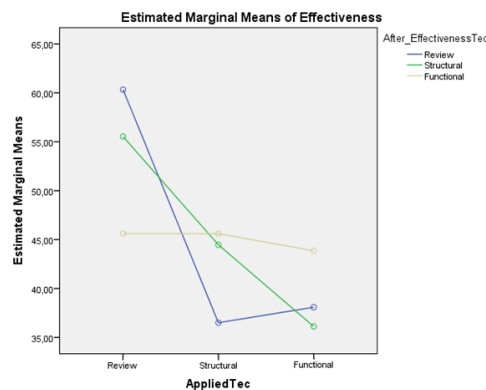


Figura 4.4. Técnica más Efectiva (Después) * Técnica Aplicada

Finalmente, en la Tabla 4.41, se puede comprobar que de los que dicen que la técnica más efectiva es “Revisión”, aplican significativamente mejor “Revisión” que “Estructural” y también que “Funcional”. Además entre “Estructural” y “Funcional” no hay diferencias. Por lo tanto parece que en cierto grado coincide lo que responden los sujetos con lo que aplican.

Por otro lado de los que dicen que la técnica más efectiva es “Estructural”, no hay diferencias significativas cuando aplican “Revisión” y “Estructural”, ni tampoco entre “Revisión” y “Funcional”. Además entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que la técnica más efectiva es “Funcional”, no hay diferencias significativas cuando aplican “Revisión” y “Estructural”, ni tampoco entre “Revisión” y “Funcional”. Además entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

IT1(After)	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Review	Review	Structural	23,829*	6,804	,003	6,877	40,780
		Functional	22,248*	4,618	,000	10,742	33,753
	Structural	Review	-23,829*	6,804	,003	-40,780	-6,877
		Functional	-1,581	5,792	1,000	-16,010	12,848
	Functional	Review	-22,248*	4,618	,000	-33,753	-10,742
		Structural	1,581	5,792	1,000	-12,848	16,010
Structural	Review	Structural	11,067	12,730	1,000	-20,646	42,779
		Functional	19,417	8,640	,089	-2,108	40,942
	Structural	Review	-11,067	12,730	1,000	-42,779	20,646
		Functional	8,350	10,835	1,000	-18,644	35,344
	Functional	Review	-19,417	8,640	,089	-40,942	2,108
		Structural	-8,350	10,835	1,000	-35,344	18,644
Functional	Review	Structural	,005	7,153	1,000	-17,816	17,826
		Functional	1,758	4,855	1,000	-10,338	13,854
	Structural	Review	-,005	7,153	1,000	-17,826	17,816
		Functional	1,753	6,089	1,000	-13,417	16,922
	Functional	Review	-1,758	4,855	1,000	-13,854	10,338
		Structural	-1,753	6,089	1,000	-16,922	13,417

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.41. Comparación por pares

4.1.1.2.4. Discusión

Para facilitar la lectura, en la tabla resumen, únicamente aparecen los datos significativos. En la Tabla 4.42, se observa que en la iteración 1, que los sujetos encuentran más en “Funcional” que en “Revisión”. Sin embargo, en la iteración 3, encuentran más en “Revisión” que en “Funcional”.

En las iteraciones 1y 2 no influye la intuición que se tiene sobre la técnica, ni la interacción técnica elegida con técnica aplicada. Además en la iteración 2, tampoco influye ni la técnica aplicada.

Por otro lado, en la iteración 3, con respecto a la interacción técnica elegida con técnica aplicada, se observa que cuando los sujetos responden “Revisión”, encuentran más faltas con la misma que con “Estructural” y “Funcional”. En el resto de los casos las técnicas se comportan igual.

Debido a que la iteración 1 y 3 aunque hay técnicas que detectan más faltas, no es la misma en los dos casos, por lo tanto, parece que realmente no es la técnica la que determina este comportamiento. Existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que podrían estar influyendo al aplicar las técnicas. Finalmente, parece que la intuición sirve para deducir lo que ha pasado, sin embargo no sirve para predecir, hay que tener en cuenta que únicamente los que intuyen “Revisión” aciertan, así que, la intuición obtenida no es muy fiable, no se podría considerar que la intuición sirva para predecir el comportamiento de las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,037	F>R	,704	N/A	,501	N/A
		R=E				
		E=F				
Iteración 2	,220	N/A	,283	N/A	,167	N/A
Iteración 3	,001	R>F	,998	N/A	,047	R R>(E=F)
		R=E				E R=E=F
		E=F				F R=E=F

Tabla 4.42 Resumen

4.1.2. ¿Qué técnica has aplicado mejor?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y la técnica que los sujetos consideran que han aplicado mejor. Esta pregunta aparece en las tres iteraciones y se realizó después de aplicar las técnicas.

4.1.2.1. Respuestas Después

4.1.2.1.1. Iteración 1

En la Tabla 4.43, se observa que para las técnicas “Estructural” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

IT2	AppliedTec	Mean	Std. Desviation	N
Review	Review	61,9033	21,81852	3
	Structural	28,5700	49,48469	3
	Functional	90,4733	8,25034	3
Structural	Review	51,7863	21,51088	8
	Structural	81,8425	6,47899	8
	Functional	81,5463	10,92748	8
Functional	Review	48,5720	27,42472	15
	Structural	73,1733	31,36062	15
	Functional	83,9667	13,42199	15

Tabla 4.43. Estadísticos Descriptivos

En la Tabla 4.44, se observa que no se puede asumir que las matrices de covarianzas son iguales ($\text{sig} \leq 0.05$).

Box's M ^a	25,318
F	3,438
df1	6
df2	1303,482
Sig.	,002

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

Tabla 4.44. Test de igualdad de matrices de covarianzas

Aunque no se puede evaluar el test de Mauchly, en la Tabla 4.45, se puede observar que tampoco era posible asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados valor de Huyn-Feldt.

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,740	6,615	2	,037	,794	,917	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.45. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 4.46 como en la Tabla 4.47, se observa que el porcentaje de faltas encontradas no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que creen que aplicaron mejor, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,572	14,718b	2,000	22,000	,000
	Wilks' Lambda	,428	14,718b	2,000	22,000	,000
	Hotelling's Trace	1,338	14,718b	2,000	22,000	,000
	Roy's Largest Root	1,338	14,718b	2,000	22,000	,000
AppliedTec * IT2	Pillai's Trace	,292	1,965	4,000	46,000	,116
	Wilks' Lambda	,711	2,043b	4,000	44,000	,105
	Hotelling's Trace	,402	2,109	4,000	42,000	,097
	Roy's Largest Root	,391	4,491c	2,000	23,000	,023

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 4.46. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	9194,195	1,588	5790,929	7,666	,003
	Huynh-Feldt	9194,195	1,834	5011,950	7,666	,002
	Lower-bound	9194,195	1,000	9194,195	7,666	,011
AppliedTec * IT2	Greenhouse-Geisser	6181,363	3,175	1946,654	2,577	,066
	Huynh-Feldt	6181,363	3,669	1684,796	2,577	,056
	Lower-bound	6181,363	2,000	3090,681	2,577	,098
Error(AppliedTec)	Greenhouse-Geisser	27584,765	36,517	755,398		
	Huynh-Feldt	27584,765	42,192	653,784		
	Lower-bound	27584,765	23,000	1199,338		

Tabla 4.47. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} \leq 0.05$), sin embargo, Roy's Largest Root es significativo en la interacción ($\text{sig} \leq 0.05$).

En la Tabla 4.48, se observa que las opiniones de la técnica mejor aplicada no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	229970,643	1	229970,643	551,533	,000
IT2	852,092	2	426,046	1,022	,376
Error	9590,224	23	416,966		

Tabla 4.48. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- La intuición sirve para deducir lo que ha sucedido en realidad. Por lo tanto, podría garantizar que una técnica se haya aplicado mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 4.49, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y entre los que aplicaron “Estructural” y “Revisión”, aunque no hay mucha diferencia, la técnica con menor media es “Estructural”. Esto se puede observar gráficamente en la Figura 4.5.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	54,087	6,110	41,447	66,727
Structural	61,195	6,935	46,850	75,540
Functional	85,329	2,977	79,171	91,487

Tabla 4.49. Medias estimadas

Finalmente, en la Tabla 4.50, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

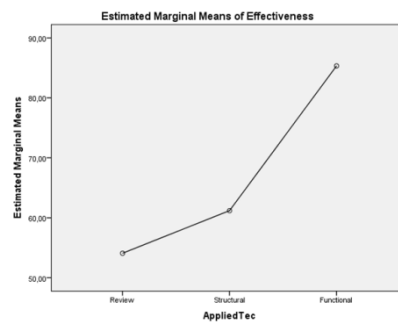


Figura 4.5. Técnica Aplicada

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	-7,108	10,262	1,000	-33,604	19,388
	Funcional	-31,242 [*]	6,946	,000	-49,176	-13,307
Structural	Review	7,108	10,262	1,000	-19,388	33,604
	Funcional	-24,133 [*]	7,506	,012	-43,513	-4,754
Funcional	Review	31,242 [*]	6,946	,000	13,307	49,176
	Structural	24,133 [*]	7,506	,012	4,754	43,513

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.50. Comparación por pares

Con respecto a la interacción técnica elegida con técnica aplicada, en la Tabla 4.51 se puede observar que únicamente cuando responden “Funcional” coincide la técnica que han aplicado mejor con la que han encontrado más faltas. En el resto de los casos los sujetos aplican mejor “Funcional”.

IT2	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	61,903	14,606	31,688	92,119
	Structural	28,570	16,577	-5,721	62,861
	Funcional	90,473	7,116	75,753	105,194
Structural	Review	51,786	8,944	33,283	70,289
	Structural	81,842	10,151	60,843	102,842
	Funcional	81,546	4,358	72,532	90,561
Funcional	Review	48,572	6,532	35,059	62,085
	Structural	73,173	7,413	57,838	88,509
	Funcional	83,967	3,182	77,383	90,550

Tabla 4.51. Medias estimadas

En la Figura 4.6, se observa que los sujetos que dicen que la técnica que han aplicado mejor es “Revisión” aplican bastante peor “Estructural” y entre “Revisión” y “Funcional” no hay mucha diferencia, sin embargo la que realmente han aplicado mejor es “Funcional”. De los que respondieron “Estructural”, aunque no hay mucha diferencia con “Funcional”, efectivamente aplican mejor “Estructural”, y aplican bastante peor “Revisión”. Finalmente de los que respondieron “Funcional”, efectivamente aplican mejor la misma, sin embargo, no hay mucha diferencia con “Estructural” y aplican bastante peor “Revisión”

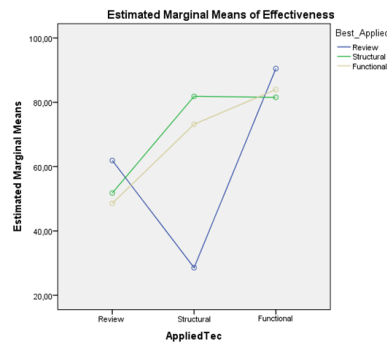


Figura 4.6. Técnica Aplicada * Técnica mejor aplicada

Finalmente en la Tabla 4.52, se puede comprobar que de los que dicen que la técnica mejor aplicada es “Revisión”, aplican significativamente mejor “Funcional” que “Estructural”, sin embargo, entre “Revisión” y “Funcional” no hay diferencias. Y entre “Revisión” y “Estructural” tampoco hay diferencias significativas. Por lo tanto no coincide lo que responden los sujetos con lo que aplican.

Por otro lado, los que dicen que la técnica mejor aplicada es “Estructural”, aplican mejor “Funcional” que “Revisión”, y no hay diferencias entre “Revisión” y “Estructural”, ni tampoco entre “Estructural” y “Funcional”. Por lo tanto no coincide lo que responden los sujetos con lo que aplican

Finalmente de los que dicen que la técnica mejor aplicada es “Funcional”, aplican significativamente mejor “Funcional” que “Revisión”, sin embargo, no hay diferencias significativas entre “Revisión” y “Estructural”, ni tampoco entre “Estructural” y “Funcional”. Por lo tanto parece que coincide lo que responden los sujetos con lo que aplican.

IT2	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Review	Review	Structural	33,333	24,531	,562	-30,005	96,672
		Functional	-28,570	16,604	,296	-71,441	14,301
	Structural	Review	-33,333	24,531	,562	-96,672	30,005
		Functional	-61,903*	17,942	,007	-108,229	-15,578
	Functional	Review	28,570	16,604	,296	-14,301	71,441
		Structural	61,903*	17,942	,007	15,578	108,229
Structural	Review	Structural	-30,056	15,022	,172	-68,843	8,730
		Functional	-29,760*	10,168	,023	-56,013	-3,507
	Structural	Review	30,056	15,022	,172	-8,730	68,843
		Functional	,296	10,987	1,000	-28,072	28,665
	Functional	Review	29,760*	10,168	,023	3,507	56,013
		Structural	-,296	10,987	1,000	-28,665	28,072
Functional	Review	Structural	-24,601	10,970	,105	-52,927	3,724
		Functional	-35,395*	7,425	,000	-54,567	-16,222
	Structural	Review	24,601	10,970	,105	-3,724	52,927
		Functional	-10,793	8,024	,575	-31,511	9,924
	Functional	Review	35,395*	7,425	,000	16,222	54,567
		Structural	10,793	8,024	,575	-9,924	31,511

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.52. Comparación por pares

4.1.2.1.2. Iteración 2

En la Tabla 4.53, se observa que para el caso de las técnicas “Revisión” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si la coincidencia es significativa.

IT2	AppliedTec	Mean	Std. Desviation	N
Review	Review	60,0000	16,10153	10
	Structural	38,3333	17,65583	10
	Functional	48,3330	14,59363	10
Structural	Review	58,3333	22,97341	6
	Structural	52,7778	16,38653	6
	Functional	33,3333	25,81989	6
Functional	Review	47,7273	25,35226	22
	Structural	47,7273	17,28646	22
	Functional	50,0000	15,43033	22

Tabla 4.53. Estadísticos Descriptivos

En la Tabla 4.54, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	18,124
F	1,224
df1	12
df2	1089,624
Sig.	,261

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

Tabla 4.54. Test de igualdad de matrices de covarianzas

En la Tabla 4.55, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,963	1,296	2	,523	,964	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.55. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 4.56, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que creen que aplicaron mejor, se puede observar que la influencia es significativa ($\text{sig} \leq 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	2109,708	2	1054,854	3,004	,056
AppliedTec * IT2	Sphericity Assumed	3525,718	4	881,430	2,510	,049
Error(AppliedTec)	Sphericity Assumed	24578,708	70	351,124		

Tabla 4.56. Test univariado de efectos intra-sujetos

En la Tabla 4.57, se observa que las opiniones de la técnica mejor aplicada no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	203541,951	1	203541,951	489,133	,000
IT2	6,614	2	3,307	,008	,992
Error	14564,490	35	416,128		

Tabla 4.57. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas
- La intuición sirve para deducir lo que ha sucedido en realidad. Por lo tanto, puede garantizar que una técnica se haya aplicado mejor que otra.

Con respecto a la interacción técnica elegida con técnica aplicada, en la Tabla 4.58, se observa que los que respondieron que la técnica mejor aplicada es “Revisión” y “Funcional” coincide con la técnica que aplicaron mejor. Sin embargo, los que respondieron “Estructural” no coincide con la técnica que aplicaron mejor, dado que la técnica mejor aplicada es “Revisión”.

IT2	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	60,000	7,264	45,253	74,747
	Structural	38,333	5,457	27,255	49,412
	Functional	48,333	5,412	37,347	59,319
Structural	Review	58,333	9,378	39,295	77,372
	Structural	52,778	7,045	38,475	67,080
	Functional	33,333	6,986	19,150	47,517
Functional	Review	47,727	4,898	37,785	57,670
	Structural	47,727	3,679	40,258	55,197
	Functional	50,000	3,649	42,593	57,407

Tabla 4.58. Medias estimadas

En la Figura 4.7, se puede observar que los que respondieron “Revisión” aplican peor “Revisión”. En el caso de los que responden “Estructural” la que peor aplican es “Funcional” y las otras dos las aplican más o menos igual. Finalmente los que eligen “Funcional” aplican peor “Revisión” y “Estructural”, aunque no hay mucha diferencia entre las tres técnicas.

Además, en la Tabla 4.59, se puede comprobar que de los que dicen que la técnica mejor aplicada es “Revisión”, aplican significativamente mejor “Revisión” que “Estructural”, sin embargo entre “Revisión” y “Funcional” no hay diferencias y entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que en cierto grado coincide lo que responden con lo que aplican.

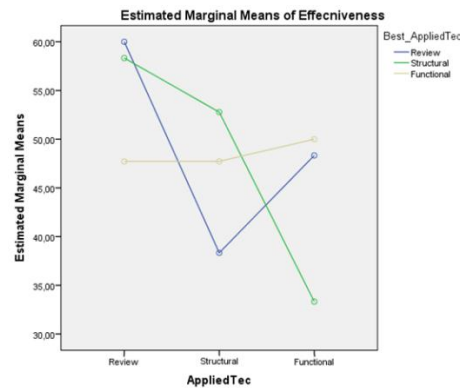


Figura 4.7. Técnica mejor Aplicada * Técnica Aplicada

Por otro lado de los que dicen que la técnica mejor aplicada es “Estructural”, no hay diferencias significativas cuando aplican “Revisión” y “Estructural”, ni tampoco entre “Revisión” y “Funcional”. Además entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que la técnica mejor aplicada es “Funcional”, no hay diferencias significativas cuando aplican “Revisión” y “Estructural”, ni tampoco entre “Revisión” y “Funcional”. Además entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

IT2	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
Review	Review	Structural	21,667*	8,611	,050	,013	43,320
		Funcional	11,667	8,917	,598	-10,756	34,090
	Structural	Review	-21,667*	8,611	,050	-43,320	-,013
		Funcional	-10,000	7,550	,582	-28,985	8,985
	Funcional	Review	-11,667	8,917	,598	-34,090	10,756
		Structural	10,000	7,550	,582	-8,985	28,985
Structural	Review	Structural	5,556	11,117	1,000	-22,399	33,510
		Funcional	25,000	11,512	,110	-3,948	53,948
	Structural	Review	-5,556	11,117	1,000	-33,510	22,399
		Funcional	19,444	9,747	,162	-5,065	43,954
	Funcional	Review	-25,000	11,512	,110	-53,948	3,948
		Structural	-19,444	9,747	,162	-43,954	5,065
Funcional	Review	Structural	1,818E-009	5,806	1,000	-14,599	14,599
		Funcional	-2,273	6,012	1,000	-17,390	12,845
	Structural	Review	-1,818E-009	5,806	1,000	-14,599	14,599
		Funcional	-2,273	5,090	1,000	-15,072	10,527
	Funcional	Review	2,273	6,012	1,000	-12,845	17,390
		Structural	2,273	5,090	1,000	-10,527	15,072

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.59. Comparación por pares

4.1.2.1.3. Iteración 3

En la Tabla 4.60 se puede observar que para el caso de las técnicas “Revisión” y “Estructural” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Funcional” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

IT2	AppliedTec	Mean	Std. Deviation	N
Review	Review	60,5421	14,93316	19
	Structural	38,5895	25,48734	19
	Functional	38,5842	13,67058	19
Structural	Review	46,6600	13,97222	5
	Structural	46,6800	21,73964	5
	Functional	33,3200	11,77336	5
Functional	Review	49,2364	24,92047	22
	Structural	42,4227	23,40374	22
	Functional	43,1818	15,14054	22

Tabla 4.60. Estadísticos Descriptivos

En la Tabla 4.61, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

En la Tabla 4.62, se puede observar que no es posible asumir esfericidad ($\text{sig} \leq 0.05$).

Box's M ^a	10,600
F	,707
df1	12
df2	579,507
Sig.	,745

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

Tabla 4.61. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,819	8,369	2	,015	,847	,919	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.62. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 4.63 como en la Tabla 4.64, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar

cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que creen que aplicaron mejor, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,216	5,791b	2,000	42,000	,006
	Wilks' Lambda	,784	5,791b	2,000	42,000	,006
	Hotelling's Trace	,276	5,791b	2,000	42,000	,006
	Roy's Largest Root	,276	5,791b	2,000	42,000	,006
AppliedTec * IT2	Pillai's Trace	,138	1,589	4,000	86,000	,185
	Wilks' Lambda	,865	1,574b	4,000	84,000	,189
	Hotelling's Trace	,152	1,558	4,000	82,000	,193
	Roy's Largest Root	,124	2,657c	2,000	43,000	,082

a. Design: Intercept + IT2

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 4.63. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3013,992	1,694	1779,256	4,079	,027
	Huynh-Feldt	3013,992	1,838	1639,784	4,079	,023
	Lower-bound	3013,992	1,000	3013,992	4,079	,050
AppliedTec * IT2	Greenhouse-Geisser	2206,757	3,388	651,360	1,493	,220
	Huynh-Feldt	2206,757	3,676	600,301	1,493	,216
	Lower-bound	2206,757	2,000	1103,378	1,493	,236
Error(AppliedTec)	Greenhouse-Geisser	31773,770	72,840	436,211		
	Huynh-Feldt	31773,770	79,036	402,017		
	Lower-bound	31773,770	43,000	738,925		

Tabla 4.64 Test univariado de efectos intra-sujetos

Todos los estadísticos de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} \leq 0.05$).

En la Tabla 4.65, se observa que las opiniones de la técnica mejor aplicada no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	178219,156	1	178219,156	385,468	,000
IT2	162,526	2	81,263	,176	,839
Error	19880,830	43	462,345		

Tabla 4.65. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- La intuición de los sujetos sobre una técnica no hace que se detecten más faltas

- La intuición no sirve para deducir lo que ha sucedido en realidad. Por lo tanto, no garantiza que una técnica se haya aplicado mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 4.66, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión”. Esto se observa gráficamente en la Figura 4.8.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	52,146	3,707	44,671	59,621
Structural	42,564	4,396	33,699	51,429
Functional	38,362	2,593	33,132	43,592

Tabla 4.66. Medias estimadas

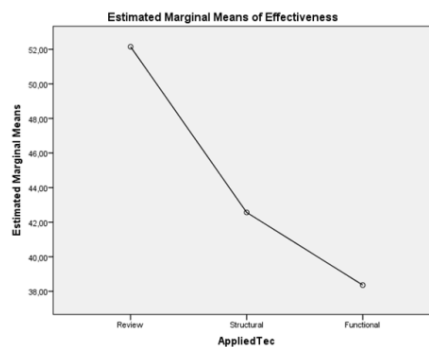


Figura 4.8. Técnica Aplicada

Finalmente, en la Tabla 4.67, se puede comprobar que la técnica “Revisión” es significativamente mejor que la “Funcional”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	9,582	5,841	,325	-4,968	24,133
	Functional	13,784*	4,033	,004	3,736	23,832
Structural	Review	-9,582	5,841	,325	-24,133	4,968
	Functional	4,202	4,800	1,000	-7,756	16,161
Functional	Review	-13,784*	4,033	,004	-23,832	-3,736
	Structural	-4,202	4,800	1,000	-16,161	7,756

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.67. Medias estimadas

4.1.2.1.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 4.68, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas en la misma que en “Revisión”. Con respecto a la interacción, cuando los sujetos responden “Revisión”, encuentran más faltas en “Funcional” que en

“Estructural”, cuando responden “Estructural” y “Funcional” encuentran más faltas en “Funcional” que en “Revisión”.

En la iteración 3, cuando los sujetos aplican “Revisión” encuentran más faltas la misma que en “Funcional”. Además no influye la intuición que se tiene sobre la técnica, ni la interacción técnica elegida con técnica aplicada.

Por otro lado, en la iteración 2, no influye ni la técnica aplicada ni la técnica elegida, sin embargo si influye la interacción técnica elegida con técnica aplicada, se observa que cuando los sujetos responden “Revisión”, encuentran más faltas con la misma que con “Estructural” pero no que en “Funcional”. Además en el resto de los casos las técnicas se comportan de igual manera.

Debido a que la iteración 1 y 3 aunque hay técnicas que detectan más faltas, no es la misma en los dos casos, por lo tanto, parece que realmente no es la técnica la que determina este comportamiento. Existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que pueden estar influyendo al aplicar las técnicas. Finalmente, parece que en cierta manera la intuición sirve para deducir lo que ha pasado, sin embargo no sirve para predecir, sin embargo, hay que tener en cuenta que los resultados no son nada determinantes, así que no se podría considerar que la intuición sirva para predecir el comportamiento de las técnicas de evaluación de código.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,000	F>R	,376	N/A	,116	R
						F>E
		R=E				R=E
						R=F
		E=F				E
						F>R
Iteración 2	,056	N/A	,992	N/A	,049	R
						R>E
						R=F
						E=F
						E
						R=E=F
Iteración 3	,006	R>F	,839	N/A	,185	F
		R=E				R=E=F
		E=F				N/A

Tabla 4.68 Resumen

4.2. Intuiciones sobre los Programas

4.2.1. ¿Con qué programa has detectado más faltas?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa en el que los sujetos consideran que han encontrado más faltas. Esta pregunta aparece en la segunda y tercera iteración; además únicamente se preguntó a los sujetos después del estudio empírico.

4.2.1.1. Resultados Después

4.2.1.1.1. Iteración 2

En la Tabla 4.69, se puede observar que para el caso de los programas “Nametbl” y “Ntree” coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de “Cmdline” no coincide ya que el programa en el que han encontrado más faltas es “Nametbl”. A continuación, se va a analizar si esta coincidencia es significativa.

IP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	51,0413	18,72706	16
	Nametbl	54,1669	15,51767	16
	Ntree	49,9994	19,24539	16
Nametbl	Cmdline	33,3323	16,66667	13
	Nametbl	55,1292	12,52030	13
	Ntree	50,0000	23,56964	13
Ntree	Cmdline	25,0013	19,92028	8
	Nametbl	47,9162	16,52049	8
	Ntree	58,3337	17,81898	8

Tabla 4.69. Estadísticos Descriptivos

En la Tabla 4.70, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	12,165
F	,864
df1	12
df2	2586,129
Sig.	,583

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IP1

Within Subjects Design: AppliedProg

Tabla 4.70. Test de igualdad de matrices de covarianzas

En la Tabla 4.71, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,979	,697	2	,706	,980	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.71. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 4.72, se observa que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción

del programa que los sujetos aplicaron y el programa que consideran el más efectivo, se puede observar que la influencia es significativa ($\text{sig} \leq 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	5907,961	2	2953,981	9,875	,000
AppliedProg * IP1	Sphericity Assumed	3789,992	4	947,498	3,168	,019
Error(AppliedProg)	Sphericity Assumed	20340,393	68	299,123		

Tabla 4.72. Test univariado de efectos intra-sujetos

En la Tabla 4.73, se observa que las opiniones del programa más efectivo no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	227611,566	1	227611,566	600,862	,000
IP1	1235,992	2	617,996	1,631	,211
Error	12879,480	34	378,808		

Tabla 4.73. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- La intuición de los sujetos sobre un programa no hace que se detecten más faltas
- La intuición sirve para deducir lo que ha sucedido en realidad. Por lo tanto, puede garantizar que un programa se haya aplicado mejor que otro.

Con respecto al programa aplicado, en la Tabla 4.74, se puede observar que la media más alta la tienen los sujetos que aplicaron el programa “Ntree” y la más baja los que aplicaron “Cmdline”. Esto se puede observar gráficamente en la Figura 4.9.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	36,458	3,135	30,088	42,829
Nametbl	52,404	2,529	47,264	57,544
Ntree	52,778	3,533	45,598	59,958

Tabla 4.74. Medias estimadas

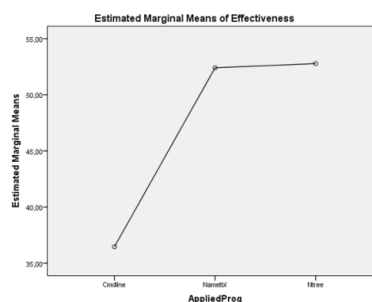


Figura 4.9. Programa Aplicado

Finalmente, en la Tabla 4.75, se puede comprobar que el programa “Nametbl” es significativamente mejor que “Cmdline”. Además se puede comprobar que el programa “Ntree” es

significativamente mejor que “Cmdline”. Finalmente, no existen diferencias significativas entre “Nametbl” y “Ntree”.

(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-15,946*	3,891	,001	-25,745	-6,146
	Ntree	-16,319*	4,408	,002	-27,419	-5,220
Nametbl	Cmdline	15,946*	3,891	,001	6,146	25,745
	Ntree	-,374	4,261	1,000	-11,105	10,358
Ntree	Cmdline	16,319*	4,408	,002	5,220	27,419
	Nametbl	,374	4,261	1,000	-10,358	11,105

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.75. Comparación por pares

Con respecto la interacción del programa elegido con el programa aplicado, en la Tabla 4.76, se puede observar que de los que respondieron que el programa en el que encontraron más faltas era “Cmdline” encontraron más faltas en “Nametbl”. De los que respondieron que el programa en el que encontraron más faltas era “Nametbl” coincide con el programa en el que encontraron más faltas. Finalmente, de los que respondieron que el programa en el que encontraron más faltas era “Ntree” coincide con el programa en el que encontraron más faltas.

IP1	AppliedProg	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Cmdline	Cmdline	51,041	4,572	41,750	60,333
	Nametbl	54,167	3,689	46,670	61,664
	Ntree	49,999	5,153	39,527	60,471
Nametbl	Cmdline	33,332	5,072	23,024	43,640
	Nametbl	55,129	4,093	46,812	63,447
	Ntree	50,000	5,717	38,382	61,618
Ntree	Cmdline	25,001	6,466	11,861	38,141
	Nametbl	47,916	5,217	37,314	58,519
	Ntree	58,334	7,287	43,524	73,143

Tabla 4.76. Medias estimadas

En la Figura 4.10, se puede observar que cuando responden “Nametbl” y “Ntree” coincide lo con el programa en el que encontraron más faltas. Cuando responden “Cmdline” no coincide pues el programa en el que encuentran más faltas es “Nametbl”.

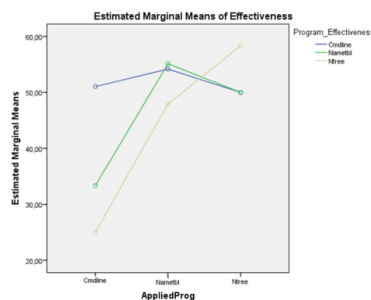


Figura 4.10. Programa más Efectivo * Programa Aplicado

Finalmente, en la Tabla 4.77, se puede comprobar que de los que dicen que el programa en el que detectan más faltas es “Cmdline” no hay diferencias significativas en las faltas encontradas con “Cmdline” y “Nmetbl”, ni tampoco entre “Cmdline” y “Ntree”. Además entre “Nmetbl” y “Ntree” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

Por otro lado de los que dicen que el programa en el que detectan más faltas es “Nmetbl”, este es significativamente mejor que “Cmdline”, entre “Nmetbl” y “Ntree” no hay diferencias significativas. Además entre “Nmetbl” y “Ntree” tampoco hay diferencias significativas. Por lo tanto parece que en cierta manera coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que el programa en el que detectan más faltas es “Ntree”, este es significativamente mejor que “Cmdline”, además se puede comprobar que “Nmetbl” también es significativamente mejor que “Cmdline”. Finalmente entre “Nmetbl” y “Ntree” no hay diferencias significativas. Por lo tanto parece que en cierta manera coincide lo que responden los sujetos con lo que aplican.

IP1	(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
Cmdline	Cmdline	Nmetbl	-3,126	5,676	1,000	-17,419	11,167
		Ntree	1,042	6,429	1,000	-15,147	17,231
	Nmetbl	Cmdline	3,126	5,676	1,000	-11,167	17,419
		Ntree	4,168	6,215	1,000	-11,484	19,819
	Ntree	Cmdline	-1,042	6,429	1,000	-17,231	15,147
		Nmetbl	-4,168	6,215	1,000	-19,819	11,484
Nmetbl	Cmdline	Nmetbl	-21,797*	6,297	,004	-37,654	-5,940
		Ntree	-16,668	7,132	,076	-34,628	1,292
	Nmetbl	Cmdline	21,797*	6,297	,004	5,940	37,654
		Ntree	5,129	6,895	1,000	-12,235	22,493
	Ntree	Cmdline	16,668	7,132	,076	-1,292	34,628
		Nmetbl	-5,129	6,895	1,000	-22,493	12,235
Ntree	Cmdline	Nmetbl	-22,915*	8,027	,022	-43,128	-2,702
		Ntree	-33,332*	9,091	,002	-56,227	-10,438
	Nmetbl	Cmdline	22,915*	8,027	,022	2,702	43,128
		Ntree	-10,418	8,790	,732	-32,553	11,718
	Ntree	Cmdline	33,332*	9,091	,002	10,438	56,227
		Nmetbl	10,418	8,790	,732	-11,718	32,553

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.77. Comparación por pares

4.2.1.1.2. Iteración 3

En la Tabla 4.78, se puede observar que para el caso del programa “Nmetbl” coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de “Cmdline” y “Ntree” no coincide ya que en ambos casos el programa en el que han encontrado más faltas es “Nmetbl”. A continuación, se va a analizar si esta coincidencia es significativa.

IP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	47,2222	14,29315	18
	Nametbl	52,7778	21,57916	18
	Ntree	44,4433	22,86676	18
Nametbl	Cmdline	34,4433	11,72796	15
	Nametbl	55,5560	16,26703	15
	Ntree	35,5553	19,78725	15
Ntree	Cmdline	31,9442	24,05498	12
	Nametbl	52,7775	21,12178	12
	Ntree	48,6117	24,05672	12

Tabla 4.78. Estadísticos Descriptivos

En la Tabla 4.79, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	10,774
F	,800
df1	12
df2	6596,025
Sig.	,651

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IP1

Within Subjects Design: AppliedProg

Tabla 4.79. Test de igualdad de matrices de covarianzas

En la Tabla 4.80, se observa que también es posible asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,962	1,606	2	,448	,963	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.80. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 4.81, se observa que el porcentaje de faltas encontradas no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción del programa aplicado y el programa considerado el más efectivo, la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	5736,855	2	2868,428	8,087	,001
AppliedProg * Effectiveness	Sphericity Assumed	2449,905	4	612,476	1,727	,152
Error(AppliedProg)	Sphericity Assumed	29793,767	84	354,688		

Tabla 4.81. Test univariado de efectos intra-sujetos

En la Tabla 4.82, se observa la elección del programa más efectivo no es significativa ($\text{sig}>0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	263799,186	1	263799,186	576,017	,000
IP1	991,732	2	495,866	1,083	,348
Error	19234,795	42	457,971		

Tabla 4.82. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- La intuición de los sujetos sobre un programa no hace que se detecten más faltas
- La intuición no sirve para deducir lo que ha sucedido en realidad. Por lo tanto, no garantiza que un programa se haya aplicado mejor que otro.

Para el programa aplicado, en la Tabla 4.83, se observa que la media más alta y más baja la tienen los del programa “Nametbl” y “Cmdline”, respectivamente. Esto se observa en la Figura 4.11.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	37,870	2,529	32,766	42,974
Nametbl	53,704	2,998	47,654	59,754
Ntree	42,870	3,358	36,093	49,647

Tabla 4.83. Medias estimadas

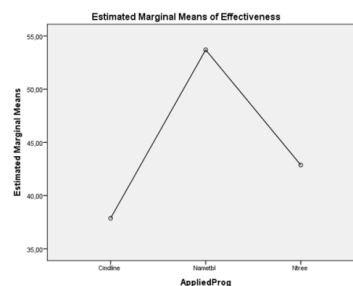


Figura 4.11. Programa Aplicado

Finalmente, en la Tabla 4.84, se puede comprobar que el programa “Nametbl” es significativamente mejor que el programa “Cmdline”. Además se puede comprobar que no existen diferencias significativas entre el programa “Ntree” y el programa “Cmdline”. Finalmente, no existen diferencias significativas entre el programa “Nametbl” y el programa “Ntree”.

(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-15,834*	3,948	,001	-25,679	-5,989
	Ntree	-5,000	3,716	,557	-14,268	4,267
Nametbl	Cmdline	15,834*	3,948	,001	5,989	25,679
	Ntree	10,834	4,383	,053	-,095	21,762
Ntree	Cmdline	5,000	3,716	,557	-4,267	14,268
	Nametbl	-10,834	4,383	,053	-21,762	,095

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.84. Comparación por pares

4.2.1.1.3. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 4.85, se observa que en la iteración 2, cuando los sujetos aplican el programa “Cmdline” encuentran más faltas en “Nametbl” y “Ntree” que en “Cmdline”. Sin embargo, en la iteración 3, cuando aplican “Cmdline” encuentran más faltas en “Nametbl” que en “Cmdline”.

En las iteraciones 2 y 3 no influye la intuición que se tiene sobre el programa. Además en la iteración 3 tampoco influye la interacción programa elegido con programa aplicado.

Por otro lado, en la iteración 2, con respecto a la interacción programa elegido con programa aplicado, se observa que cuando los sujetos responden “Nametbl”, encuentran más faltas en el mismo que en “Cmdline” pero no que en “Ntree”. Además en el resto de los casos los programas se comportan igual. Por otro lado, cuando responden “Ntree” encuentran menos faltas en “Cmdline”.

Debido a que la iteración 2 y 3 el programa en el que detectan más faltas, es el mismo en los dos casos, realmente algunos programas permiten encontrar más faltas que otros. Finalmente, parece que la intuición sirve para deducir lo que ha pasado, sin embargo, no se podría considerar que la intuición sirva para predecir el comportamiento de las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares	
Iteración 2	,000	(na=nt)>cm	,211	N/A	,019	cm	cm=na=nt
						na	na>cm
							nt=cm
							na=nt
Iteración 3	,001	na>cm	,348	N/A	,152	nt	(na=nt)>cm
		nt=cm				N/A	
		na=nt					

Tabla 4.85 Resumen

4.2.2. ¿Con qué programa has detectado menos faltas?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa en el que los sujetos consideran que han encontrado menos faltas. Esta pregunta aparece únicamente en la tercera iteración; además únicamente se preguntó a los sujetos después del estudio empírico.

4.2.2.1. Respuestas Después

4.2.2.1.1. Iteración 3

En la Tabla 4.86, se puede observar que en los tres programas coincide el programa que han elegido los sujetos con el programa en el que han encontrado menos faltas. Además se esperaba esta coincidiera.

En la Tabla 4.87, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

IP2	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	23,3327	15,16947	15
	Nametbl	56,6673	16,42897	15
	Ntree	41,1107	17,66920	15
Nametbl	Cmdline	48,8893	18,32954	15
	Nametbl	43,3333	20,70139	15
	Ntree	46,6660	26,87469	15
Ntree	Cmdline	41,6656	12,17268	16
	Nametbl	61,4581	16,90791	16
	Ntree	40,6250	21,91596	16

Tabla 4.86. Estadísticos Descriptivos

Box's M ^a	21,733
F	1,626
df1	12
df2	8879,999
Sig.	,077

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + IP2

Within Subjects Design: AppliedProg

Tabla 4.87. Test de igualdad de matrices de covarianzas

En la Tabla 4.88, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,953	2,040	2	,361	,955	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + IP2

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 4.88. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 4.89, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran el menos efectivo, se puede observar que la influencia es significativa ($\text{sig} \leq 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	6070,548	2	3035,274	9,664	,000
AppliedProg * IP2	Sphericity Assumed	6818,805	4	1704,701	5,428	,001
Error(AppliedProg)	Sphericity Assumed	27010,674	86	314,078		

Tabla 4.89. Test univariado de efectos intra-sujetos

En la Tabla 4.90, se observa que las opiniones del programa menos efectivo no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	277468,110	1	277468,110	631,589	,000
IP2	1447,664	2	723,832	1,648	,204
Error	18890,640	43	439,317		

Tabla 4.90. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan menos faltas
- La intuición de los sujetos sobre un programa no hace que se detecten más faltas
- La intuición sirve para deducir lo que ha sucedido en realidad. Por lo tanto, puede garantizar que un programa se haya aplicado mejor que otro.

Con respecto al programa aplicado, en la Tabla 4.91, se puede observar que la media más baja tienen los sujetos que aplicaron el programa “Cmdline” y la más alta los que aplicaron el programa “Nametbl”. Esto se puede comprobar gráficamente en la Figura 4.12.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	37,963	2,266	33,393	42,533
Nametbl	53,820	2,668	48,439	59,200
Ntree	42,801	3,313	36,120	49,481

Tabla 4.91. Medias estimadas

Finalmente, en la Tabla 4.92, se puede comprobar que el programa “Cmdline” es significativamente peor que “Nametbl”. Además no hay diferencias significativas entre el programa “Cmdline” y “Ntree”. Finalmente, el programa “Ntree” es significativamente peor que “Nametbl”.

(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-15,857*	3,349	,000	-24,199	-7,515
	Ntree	-4,838	3,666	,582	-13,971	4,295
Nametbl	Cmdline	15,857*	3,349	,000	7,515	24,199
	Ntree	11,019*	4,044	,028	,945	21,093
Ntree	Cmdline	4,838	3,666	,582	-4,295	13,971
	Nametbl	-11,019*	4,044	,028	-21,093	-,945

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.92. Comparación por pares

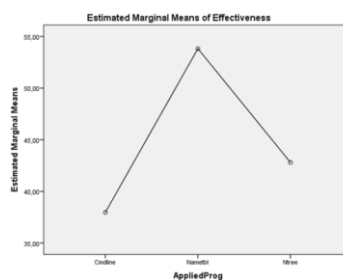


Figura 4.12. Programa Aplicado

Con respecto a la interacción programa elegido con programa aplicado, en la Tabla 4.93, se puede observar que en los tres programas coincide el programa que han elegido los sujetos con el programa en el que han encontrado menos faltas. Además se esperaba esta coincidiera.

IP2	AppliedProg	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Cmdline	Cmdline	23,333	3,966	15,333	31,332
	Nametbl	56,667	4,670	47,249	66,085
	Ntree	41,111	5,799	29,417	52,805
Nametbl	Cmdline	48,889	3,966	40,890	56,889
	Nametbl	43,333	4,670	33,915	52,751
	Ntree	46,666	5,799	34,972	58,360
Ntree	Cmdline	41,666	3,841	33,920	49,411
	Nametbl	61,458	4,522	52,339	70,577
	Ntree	40,625	5,614	29,302	51,948

Tabla 4.93. Medias estimadas

En la Figura 4.13, se puede observar que los que respondieron “Cmdline” coincide con el programa en el que encontraron menos faltas. En el que encuentran más faltas es “Nametbl”. En el caso de los que responden “Nametbl” coincide con el programa en el que encontraron menos faltas y entre “Cmdline” y “Ntree” aunque no hay mucha diferencia en el que encuentran más faltas es “Cmdline”. Finalmente los que eligen “Ntree” coincide con el programa en el que encontraron menos faltas y en el que encuentran más faltas es “Nametbl”.

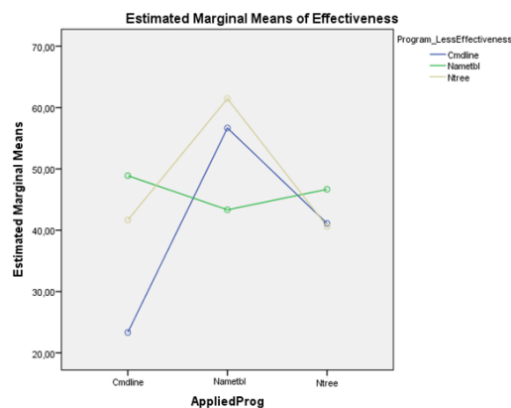


Figura 4.13. Programa menos Efectivo * Programa Aplicado

Finalmente, en la Tabla 4.94, se puede comprobar que de los que dicen que el programa en el que detectan menos faltas es “Cmdline”, “Cmdline” es significativamente peor que “Nametbl” y también significativamente peor que “Ntree”. Finalmente, entre “Nametbl” y “Ntree” no hay diferencias significativas. Por lo tanto parece que coincide lo que responden con la realidad.

Por otro lado de los que dicen que el programa en el que detectan menos faltas es “Nametbl”, no hay diferencias significativas con “Cmdline”, entre “Nametbl” y “Ntree” tampoco hay diferencias significativas. Finalmente entre “Nametbl” y “Ntree” tampoco hay diferencias significativas. Por lo tanto parece que no coincide lo que responden los sujetos con lo que sucede en realidad.

Finalmente de los que dicen que el programa en el que detectan menos faltas es “Ntree”, éste es significativamente peor que “Nametbl”, por otro lado, el programa “Cmdline” es significativamente peor que “Nametbl”. Finalmente entre “Cmdline” y “Ntree” no hay diferencias significativas. Por lo tanto parece que coincide lo que responden con lo que sucede en realidad.

IP2	(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Cmdline	Cmdline	Nametbl	-33,335*	5,861	,000	-47,937	-18,732
		Ntree	-17,778*	6,417	,025	-33,764	-1,792
	Nametbl	Cmdline	33,335*	5,861	,000	18,732	47,937
		Ntree	15,557	7,078	,100	-2,077	33,190
	Ntree	Cmdline	17,778*	6,417	,025	1,792	33,764
		Nametbl	-15,557	7,078	,100	-33,190	2,077
Nametbl	Cmdline	Nametbl	5,556	5,861	1,000	-9,046	20,158
		Ntree	2,223	6,417	1,000	-13,762	18,209
	Nametbl	Cmdline	-5,556	5,861	1,000	-20,158	9,046
		Ntree	-3,333	7,078	1,000	-20,966	14,301
	Ntree	Cmdline	-2,223	6,417	1,000	-18,209	13,762
		Nametbl	3,333	7,078	1,000	-14,301	20,966
Ntree	Cmdline	Nametbl	-19,792*	5,675	,003	-33,931	-5,654
		Ntree	1,041	6,213	1,000	-14,437	16,519
	Nametbl	Cmdline	19,792*	5,675	,003	5,654	33,931
		Ntree	20,833*	6,854	,012	3,759	37,907
	Ntree	Cmdline	-1,041	6,213	1,000	-16,519	14,437
		Nametbl	-20,833*	6,854	,012	-37,907	-3,759

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 4.94. Comparación por pares

4.2.2.1.2. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 4.93, se observa que en la iteración 3, cuando los sujetos aplican el programa “Nametbl” encuentran más faltas que en “Nametbl” y “Ntree”. Además se observa que la intuición sobre el programa no influye.

Por otro lado, con respecto a la interacción, se observa que cuando los sujetos responden “Cmdline”, encuentran menos faltas en el mismo que en “Nametbl” y en “Ntree”. Además en el resto de los casos los programas se comportan igual. Por otro lado cuando responden “Ntree” encuentran menos faltas en el mismo y en “Cmdline” que en “Nametbl”.

Finalmente se puede concluir, que efectivamente algunos programas permiten encontrar más faltas que otros. Además, la intuición sirve para deducir lo que ya ha ocurrido, pero no para predecir.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 3	,000	(cm=nt)<na	,204	N/A	,001	cm (na=nt)>cm
						na cm=na=nt
						nt (cm=nt)<na

Tabla 4.95. Resumen

CAPÍTULO 5. RESULTADOS SOBRE PREFERENCIAS

En este capítulo se van a presentar y evaluar los resultados obtenidos sobre las preferencias de los sujetos. Para ello se seguirá el método explicado en el capítulo 2. Las preguntas sobre las que se va a trabajar en este capítulo aparecen en la Tabla 5.1.

Código	Preguntas sobre Preferencias	It 1	It 2	It 3	Aparición
PT1	¿Qué técnica te gusta más?	Si	Si	Si	A/D
PT2	¿Qué técnica te ha parecido más fácil de entender?	Si	Si	Si	D
PT3	¿Qué técnica te ha parecido más fácil de aplicar?	Si	Si	Si	A/D
PP1	¿Qué programa es más fácil?	Si	Si	Si	D
PP2	¿Qué programa tiene las faltas más difíciles?	Si	Si	Si	D
PP3	¿Qué programa has entendido mejor?	No	Si	Si	D
PP1	¿Qué programa tiene las faltas más fáciles?	No	No	Si	D

Tabla 5.1. Preguntas sobre Preferencia

5.1. Preferencias sobre las Técnicas

5.1.1. ¿Qué técnica te gusta más?

Dado que esta pregunta se realiza tanto antes como después de la realización del estudio empírico, se va a evaluar si ha habido o no cambios de opinión con respecto a la técnica que más les gusta a los sujetos, es decir, cual es su técnica preferida. Esta pregunta aparece en las tres iteraciones.

5.1.1.1. Distribución de la Técnica

Primero se va a analizar la distribución de la técnica con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución de la técnica se realizará la prueba de distribución (χ^2 de Pearson).

Se analizará primero tanto antes como después con categorías y luego las mismas sin las categorías.

5.1.1.1.1. Resultados de Antes con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.2 muestra los resultados de la categoría “Other” y la Tabla 5.3 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: más de una técnica “E/F”		1 sujeto: “Ninguna”
	1 sujeto: “Igual”		
Total	2 sujetos	0 sujetos	1 sujeto

Tabla 5.2 Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	2 sujetos
Iteración 3	3 sujetos

Tabla 5.3 Sujetos en la categoría “No answer”

En la Figura 5.1, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1, se puede observar que la mayoría de los sujetos responden la técnica “Estructural”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”; además se puede observar que hay más sujetos que no proporcionan ninguna respuesta (categoría “No answer”), que los que responden una cosa distinta (categoría “Other”).



Figura 5.1. Antes con categorías “Other” y “No answer”

5.1.1.1.2. Resultados de Después con categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.4 muestra los resultados de la categoría “Other” y la Tabla 5.5 muestra los resultados de la categoría “No answer”

En la Figura 5.2, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica

“Funcional”, sin embargo, hay un mayor porcentaje de sujetos en la categoría “Other” que los que pertenecen a la categoría “Revisión”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, sin embargo, hay un gran porcentaje de sujetos que pertenecen a la categoría “No answer”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”.

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: más de una técnica “E/F”		
	1 sujeto: más de una técnica “F/E”		
	1 sujeto: “Igual”		
Total	3 sujetos	0 sujetos	0 sujetos

Tabla 5.4. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	8 sujetos
Iteración 3	1 sujeto

Tabla 5.5 Sujetos en la categoría “No answer”



Figura 5.2. Después con categorías “Other” y “No answer”

5.1.1.1.3. Resultados de Antes sin categorías

En esta sección se han puesto a nulo los valores de los sujetos que pertenecían a las categorías: “Other” y “No answer”.

En la Figura 5.3, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica “Estructural”, seguida de la técnica “Funcional” y finalmente hay muy pocos sujetos que responden la técnica de “Revisión”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Estructural” y finalmente el resto de sujetos responden la técnica “Revisión”.



Figura 5.3. Antes sin categorías

Se puede observar que en la iteración 1 la técnica “Estructural” ocurre con mayor probabilidad, mientras que en las iteraciones 2 y 3 la técnica “Funcional” ocurre con mayor probabilidad.

5.1.1.1.3.1. Prueba de Proporciones

En la Tabla 5.6, se puede observar que para la iteración 3 las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que no hay una “preferencia definida” o “preferencia clara” hacia alguna de las técnicas. Sin embargo, en la iteración 1, las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Estructural” y “Funcional” dado que estas ocurren con mayor probabilidad que la técnica “Revisión”. Finalmente, en la iteración 2, las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	27	6,889	2	,032
Iteración 2	44	27,591	2	,000
Iteración 3	42	3,	2	,223

Tabla 5.6 Prueba de Chi-Cuadrado

En la Figura 5.4, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.7, para cada iteración.

	N observado			N esperado			Residual		
	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3
Review	3	8	10	9,0	14,7	14,0	-6,0	-6,7	-4,0
Structural	14	5	13	9,0	14,7	14,0	5,0	-9,7	-1,0
Functional	10	31	19	9,0	14,7	14,0	1,0	16,3	5,0
Total	27	44	42						

Tabla 5.7 Frecuencias para Antes

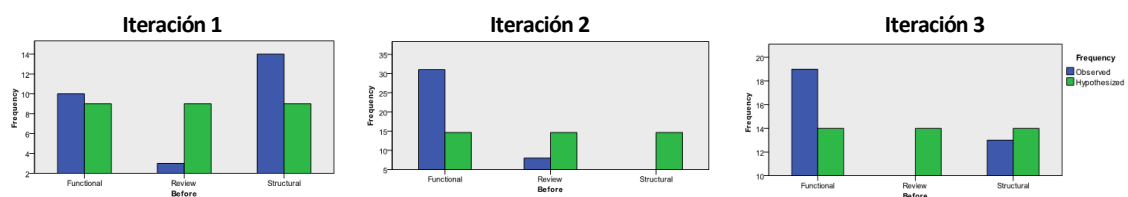


Figura 5.4. Chi-Cuadrado

5.1.1.1.4. Resultados de Después sin categorías

En esta sección se han puesto a nulo los valores de los sujetos que pertenecían a las categorías: “Other” y “No answer”.

En la Figura 5.5, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Estructural” y finalmente hay muy pocos sujetos que responden la técnica de “Revisión”. Para la iteración 2 se puede observar que la mayoría de los

sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”.



Figura 5.5. Después sin categorías

Se puede observar que en las tres iteraciones la técnica “Funcional” ocurre con mayor probabilidad.

5.1.1.1.4.1. Prueba de Proporciones

En la Tabla 5.8, se puede observar que para las tres iteraciones las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	26	17,154	2	,000
Iteración 2	38	22,474	2	,000
Iteración 3	45	17,733	2	,000

Tabla 5.8 Prueba de Chi-Cuadrado

En la Figura 5.6, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.9, para cada iteración.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	1	9	14	8,7	12,7	15,0	-7,7	-3,7	-1,0
Structural	7	3	4	8,7	12,7	15,0	-1,7	-9,7	-11,0
Functional	18	26	27	8,7	12,7	15,0	9,3	13,3	12,0
Total	26	38	45						

Tabla 5.9 Frecuencias para Después



Figura 5.6. Chi-Cuadrado

5.1.1.2. Estudio del Acuerdo Antes / Después

En la Figura 5.7, para la iteración 1, se observa que de dos personas que dicen, antes de aplicar las técnicas, que su técnica preferida es “Revisión” una cambia de opinión a “Funcional”. De catorce personas que dicen que su técnica preferida es “Estructural” siete cambian de opinión a “Funcional”. Finalmente, de diez personas dicen que su técnica preferida es “Funcional” ninguna cambia de opinión.

Para la iteración 2, de ocho personas que dicen, antes de aplicar las técnicas, que su técnica preferida es “Revisión” cinco cambian de opinión a “Funcional”. De tres personas que dicen que la técnica que más les gusta es “Estructural” una cambia de opinión a “Revisión” y otra a “Funcional”. Finalmente, de veinticinco personas que dicen que su técnica preferida es “Funcional” cuatro cambian de opinión a “Revisión” y otras dos a “Estructural”.

Para la iteración 3, de diez personas que dicen, antes de aplicar las técnicas, que la técnica más les gusta es “Revisión” cinco cambian de opinión a “Funcional”. De doce personas que dicen que la técnica que más les gusta es “Estructural” dos cambian de opinión a “Revisión” y otras seis a “Funcional”. Finalmente se observa que de diecinueve personas que dicen que la técnica que más les gusta es “Funcional” seis cambian de opinión a “Revisión”.



Figura 5.7. Cambios de Opinión

Estos resultados se pueden contrastar en la Tabla 5.10, Tabla 5.11 y Tabla 5.12 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			After			Total
			Review	Structural	Functional	
Before	Review	Count	1	0	1	2
		% of Total	3,8%	0,0%	3,8%	7,7%
	Structural	Count	0	7	7	14
		% of Total	0,0%	26,9%	26,9%	53,8%
	Functional	Count	0	0	10	10
		% of Total	0,0%	0,0%	38,5%	38,5%
Total		Count	1	7	18	26
		% of Total	3,8%	26,9%	69,2%	100,0%

Tabla 5.10 Tabla de Contingencia de la iteración 1

			After			Total
			Review	Structural	Functional	
Before	Review	Count	3	0	5	8
		% of Total	8,3%	0,0%	13,9%	22,2%
	Structural	Count	1	1	1	3
		% of Total	2,8%	2,8%	2,8%	8,3%
	Functional	Count	4	2	19	25
		% of Total	11,1%	5,6%	52,8%	69,4%
Total		Count	8	3	25	36
		% of Total	22,2%	8,3%	69,4%	100,0%

Tabla 5.11 Tabla de Contingencia de la iteración 2

			After			Total
			Review	Structural	Functional	
Before	Review	Count	5	0	5	10
		% of Total	12,2%	0,0%	12,2%	24,4%
	Structural	Count	2	4	6	12
		% of Total	4,9%	9,8%	14,6%	29,3%
	Functional	Count	6	0	13	19
		% of Total	14,6%	0,0%	31,7%	46,3%
Total		Count	13	4	24	41
		% of Total	31,7%	9,8%	58,5%	100,0%

Tabla 5.12 Tabla de Contingencia de la iteración 3

5.1.1.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 69.2%, es decir la mayoría de los sujetos no cambian de opinión. Para la iteración 2 el porcentaje es del 63.9%, es decir la mayoría de los sujetos no cambian de opinión. Finalmente para la iteración 3 el porcentaje es del 53.7%, es decir la mayoría de los sujetos no cambian de opinión.

Sin embargo en la Tabla 5.13, se puede observar que para la iteración 1, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es moderado ($>0,41 - 0,60$), se puede concluir que no existen cambios de opinión. Para la iteración 2, el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, existen cambios de opinión. Finalmente para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es discreto ($0,21 - 0,40$), se puede concluir que no existen cambios de opinión.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,475	,135	3,328	,001
Iteración 2	,217	,153	1,634	,102
Iteración 3	,256	,120	2,404	,016

Tabla 5.13. Índice de Concordancia Kappa

5.1.1.2.2. Direccionalidad del cambio de opinión

A continuación se muestran los resultados del Test de McNemar-Bowker. Según los resultados mostrados en la Tabla 5.14, en la iteración 1 y en la iteración 3, no hay simetría ($\text{sig} > 0.05$) los

cambios de una categoría a otra no se producen en ambas direcciones con igual probabilidad, es decir, los cambios tienen direccionalidad, en la Tabla 5.10, se puede observar que los sujetos tienden a cambiar hacia “Funcional” independientemente de que previamente haya respondido “Revisión” o “Estructural”. En la Tabla 5.12, se puede observar que el cambio es más probable de “Estructural” a “Revisión” que de “Revisión” a “Estructural”, por otro lado, la mayoría de sujetos cambian hacia “Funcional” independientemente de que previamente hayan contestado “Revisión” o “Estructural”, sin embargo, hay casi el mismo número de cambios de “Funcional” a “Revisión” que de “Revisión” a “Funcional”.

En la iteración 2, hay simetría ($\text{sig} > 0.05$) los cambios de una categoría a otra se producen en ambas direcciones con igual probabilidad, es decir, los cambios se producen de forma indistinta.

Iteración	Valor	gl	Sig. asintótica (bilateral)	N de casos válidos
Iteración 1	8,000	2	,018	26
Iteración 2	1,444	3	,695	36
Iteración 3	8,091	3	,044	41

Tabla 5.14. Pruebas de McNemar-Bowker

5.1.1.2.3. Resultados Prueba de Stuart-Maxwell

Finalmente se muestran los resultados del Test de Stuart-Maxwell. Según los resultados mostrados en la Tabla 5.15, para la iteración 2, la distribución de las tres técnicas es igual antes que después ($\text{sig} > 0.05$), además se esperaba este resultado puesto que también existe simetría. En el caso de la iteración 1 y la iteración 3, la distribución de las tres técnicas no es igual antes que después ($\text{sig} \leq 0.05$).

Iteración	X ²	gl	Sig.
Iteración 1	8	2	0,018
Iteración 2	0	2	1
Iteración 3	8,08	2	0,018

Tabla 5.15. Pruebas de Stuart-Maxwell

5.1.1.3. Discusión

Finalmente en la Tabla 5.16 se muestran todos los test en conjunto para comprobar si realmente existen cambios de opinión y como son estos cambios.

Para la iteración 1, se puede observar que el acuerdo es moderado. Además se puede comprobar que no hay simetría, el cambio no es simétrico ($p_{ij} \neq p_{ji}$). Finalmente se comprueba que tampoco hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica, no es el mismo. Esto se puede comprobar con el porcentaje de sujetos donde se observa que en el caso de las técnicas “Estructural” y “Funcional” el porcentaje cambia considerablemente.

Para la iteración 2, se puede observar que no hay acuerdo. Además se puede comprobar que hay simetría, es decir, cuando los sujetos cambian de opinión, el cambio es simétrico ($p_{ij} = p_{ji}$).

Finalmente se comprueba que hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica, es el mismo.

Para la iteración 3, se puede observar que el acuerdo es discreto. Además se puede comprobar que no hay simetría, el cambio no es simétrico ($p_{ij} \neq p_{ji}$). Finalmente se comprueba que tampoco hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica, no es el mismo. Esto se puede comprobar con el porcentaje de sujetos donde se observa que en el caso de las técnicas “Estructural” y “Funcional” el porcentaje cambia considerablemente.

	Acuerdo			Distribución Técnica				
	Kappa		McNemar-Bowker	Stuart-Maxwell	X ² (Antes)		X ² (Después)	
Iteración 1	,001	,475	,018	0,018	,032	R: 11,11%	,000	R: 3,85%
						E:51,85%		E:26,92%
						F: 37,04%		F: 69,23%
Iteración 2	,102		,695	1	,000	R: 18,18%	,000	R: 23,68%
						E: 11,36%		E: 7,89%
						F: 70,45%		F: 68,42%
Iteración 3	,016	,256	,044	0,018	,223	R: 23,81%	,000	R: 31,11%
						E:30,95%		E:8,89%
						F: 45,24%		F: 60,00%

Tabla 5.16. Tabla Resumen

5.1.2. ¿Qué técnica te ha parecido más fácil de entender?

En esta sección se va a evaluar la opinión de los sujetos con respecto a la técnica que consideran que es la más fácil de entender. Esta pregunta aparece en las tres iteraciones, solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

5.1.2.1. Distribución de la Técnica

Primero se va a analizar la distribución de la técnica con las categorías “Other” y “No answer”. Luego se remplazaran los datos de dichas categorías por valores nulos y además de analizar la distribución de la técnica se realizará la prueba de distribución (X^2 de Pearson).

5.1.2.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.17 muestra los resultados de la categoría “Other” y la Tabla 5.18 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: más de una técnica “F/E”		1 sujeto: más de una técnica “E/F” 1 sujeto: más de una técnica “R/E/F”
Total	1 sujeto	0 sujetos	2 sujetos

Tabla 5.17. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	10 sujetos
Iteración 2	2 sujetos
Iteración 3	3 sujetos

Tabla 5.18. Sujetos en la categoría "No answer"

En la Figura 5.8, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1, se puede observar que la mayoría de los sujetos no responden la técnica "Revisión", sin embargo, hay más sujetos en la categoría "No answer". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Revisión" y hay casi el mismo porcentaje de sujetos en las categorías "Other" y "No answer".

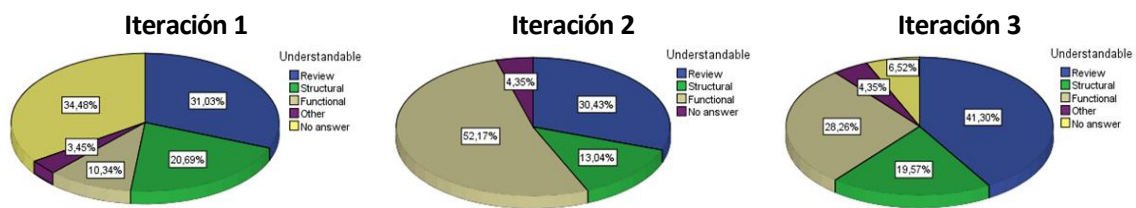


Figura 5.8. Después con categorías

5.1.2.1.2. Resultados sin categorías

En la Figura 5.9, para la iteración 1 se puede observar que la mitad de todos los sujetos responden la técnica "Revisión", seguida de la técnica "Estructural" y finalmente hay muy pocos sujetos que responden la técnica de "Funcional". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", seguida de la técnica "Revisión" y finalmente hay pocos sujetos que responden la técnica de "Estructural". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Revisión", seguida por la técnica "Funcional" y finalmente el resto de sujetos responden "Estructural".

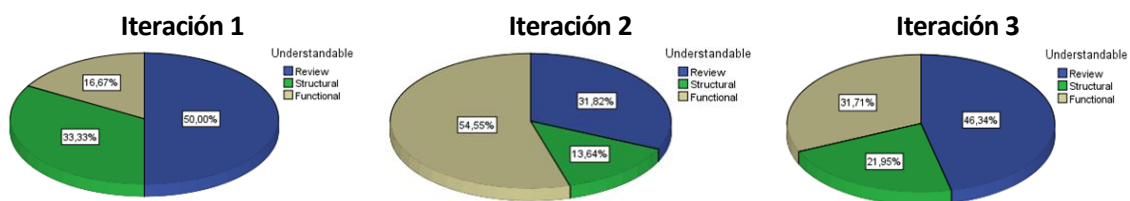


Figura 5.9. Después sin categorías

Se puede observar que en la iteración 1 y 3 es la técnica "Revisión" la que ocurre con mayor probabilidad, mientras que en la iteración 2 la técnica "Funcional" ocurre con mayor probabilidad.

5.1.2.1.2.1. Prueba de Proporciones

En la Tabla 5.19, se puede observar que en las iteraciones 1 y 3 las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una "preferencia definida" o "preferencia

clara” hacia alguna de las técnicas. Mientras que en la iteración 2 las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	15	3	2	,223
Iteración 2	44	11,091	2	,004
Iteración 3	41	3,707	2	,157

Tabla 5.19. Prueba de Chi-Cuadrado

En la Figura 5.10, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.20, para cada iteración.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	9	14	19	6,0	14,7	13,7	3,0	-,7	5,3
Structural	6	6	9	6,0	14,7	13,7	,0	-8,7	-4,7
Functional	3	24	13	6,0	14,7	13,7	-3,0	9,3	-,7
Total	18	44	41						

Tabla 5.20. Frecuencias

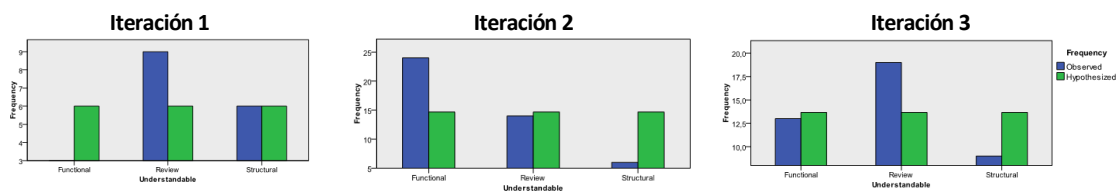


Figura 5.10. Chi-Cuadrado

5.1.3. ¿Qué técnica te ha parecido más fácil de aplicar?

Dado que esta pregunta se realiza tanto antes como después de la realización del estudio empírico, se va a evaluar si ha habido o no cambios de opinión con respecto a la técnica que los sujetos consideran que es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones.

5.1.3.1. Distribución de la Técnica

Primero se va a analizar la distribución de la técnica con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución de la técnica se realizará la prueba de distribución (χ^2 de Pearson).

Se analizará primero tanto antes como después con categorías y luego las mismas sin las categorías.

5.1.3.1.1. Resultados de Antes con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.21 muestra los resultados de la categoría “Other” y la Tabla 5.22 muestra los resultados de la categoría “No answer”

	Iteración 1	Iteración 2	Iteración 3
Respuestas			1 sujeto: "Según Programa"
Total	0 sujetos	0 sujetos	1 sujeto

Tabla 5.21. Respuestas para categoría "Other"

Iteración	Nº de sujetos
Iteración 1	1 sujeto
Iteración 2	2 sujetos
Iteración 3	3 sujetos

Tabla 5.22. Sujetos en la categoría "No answer"

En la Figura 5.11, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1, se puede observar que la mayoría de los sujetos responden la técnica "Funcional" y únicamente se ha creado la categoría "No answer". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional" y solo hay sujetos que pertenecen a la categoría "No answer". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Funcional" y se han creado tanto la categoría "No answer" como "Other".



Figura 5.11. Antes con categorías

5.1.3.1.2. Resultados de Después con Categorías "Other" y "No answer"

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.23 muestra los resultados de la categoría "No answer"

Iteración	Nº de sujetos
Iteración 1	0 sujetos
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 5.23. Sujetos en la categoría "No answer"

En la Figura 5.12, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", además no se han creado las categorías "Other" ni "No answer". Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", sin embargo, hay un gran porcentaje de sujetos que pertenecen a la categoría "No answer". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica "Funcional", además no se han creado las categorías "Other" ni "No answer".



Figura 5.12. Después con categorías

5.1.3.1.3. Resultados de Antes sin categorías

En la Figura 5.13, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Estructural” y finalmente el resto de sujetos responden la técnica de “Revisión”.

Se puede observar que en las tres iteraciones la técnica “Funcional” ocurre con mayor probabilidad.



Figura 5.13. Antes sin categorías

5.1.3.1.3.1. Prueba de Proporciones

En la Tabla 5.24, se puede observar que para la iteración 1, las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una “preferencia definida” o “preferencia clara” hacia alguna de las técnicas. Mientras que en la iteración 2 y 3 las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	28	5,214	2	,074
Iteración 2	44	27,591	2	,000
Iteración 3	42	11,228	2	,004

Tabla 5.24. Prueba de Chi-Cuadrado

En la Figura 5.14, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.25, para cada iteración.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	7	8	7	9,3	14,7	14,0	-2,3	-6,7	-7,0
Structural	6	5	11	9,3	14,7	14,0	-3,3	-9,7	-3,0
Functional	15	31	24	9,3	14,7	14,0	5,7	16,3	10,0
Total	28	44	42						

Tabla 5.25. Frecuencias para Antes



Figura 5.14. Chi-Cuadrado

5.1.3.1.4. Resultados de Después sin categorías

En la Figura 5.15, para la iteración 1 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Estructural” y finalmente el resto de sujetos que responden la técnica de “Revisión”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden la técnica “Funcional”, seguida de la técnica “Revisión” y finalmente hay muy pocos sujetos que responden la técnica de “Estructural”.



Figura 5.15. Después sin categorías

Se puede observar que en las tres iteraciones “Funcional” ocurre con mayor probabilidad.

5.1.3.1.4.1. Prueba de Proporciones

En la Tabla 5.26, se puede observar que para las tres iteraciones las tres técnicas no ocurren con la misma probabilidad ($\text{sig} \leq 0.05$), lo que significa que los sujetos prefieren “Funcional” dado que esta ocurre con mayor probabilidad que las técnicas “Revisión” y “Estructural”.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	29	13,724	2	,001
Iteración 2	38	22,474	2	,000
Iteración 3	46	28,348	2	,000

Tabla 5.26. Prueba de Chi-Cuadrado

En la Figura 5.16, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.27, para cada iteración.

	N observado			N esperado			Residual		
	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3
Review	4	9	10	9,7	12,7	15,3	-5,7	-3,7	-5,3
Structural	6	3	4	9,7	12,7	15,3	-3,7	-9,7	-11,3
Functional	19	26	32	9,7	12,7	15,3	9,3	13,3	16,7
Total	29	38	46						

Tabla 5.27. Frecuencias para Después

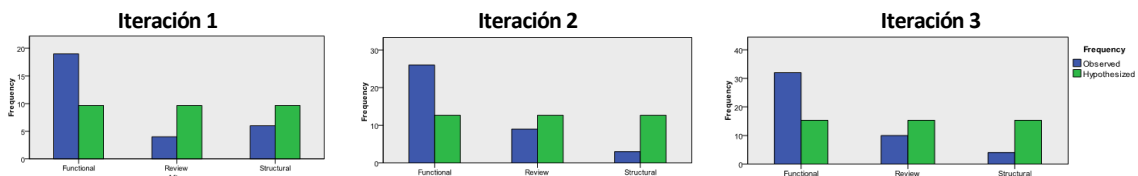


Figura 5.16. Chi-Cuadrado

5.1.3.2. Estudio del Acuerdo Antes / Después

En la Figura 5.17, para la iteración 1, se observa que de siete personas que dicen, antes de aplicar las técnicas, que la técnica más fácil de aplicar es “Revisión” dos cambian de opinión a “Funcional” y uno a “Estructural”. De seis personas que dicen que la técnica más fácil de aplicar es “Estructural” tres cambian a “Funcional”. Finalmente de quince personas que dicen que la técnica más fácil de aplicar es “Funcional” dos cambian a “Estructural”.

Para la iteración 2, de ocho personas que dicen, antes de aplicar las técnicas, que la técnica más fácil de aplicar es “Revisión” cinco cambian de opinión a “Funcional”. De tres personas que dicen que la técnica más fácil de aplicar es “Estructural” una cambia a “Revisión” y otra a “Funcional”. Finalmente se observa que de veinticinco personas que dicen que la técnica más efectiva es “Funcional” cuatro cambian a “Revisión” y otras dos a “Estructural”.

Para la iteración 3, de siete personas que dicen, antes de aplicar las técnicas, que la técnica más fácil de aplicar es “Revisión” tres cambian a “Funcional” y una a “Estructural”. De once personas que dicen que la técnica más fácil de aplicar es “Estructural” tres cambian a “Revisión” y otras siete a “Funcional”. Finalmente se observa que de veinticuatro personas que dicen que la técnica más fácil de aplicar es “Funcional” dos cambian a “Revisión” y otras tres a “Estructural”.

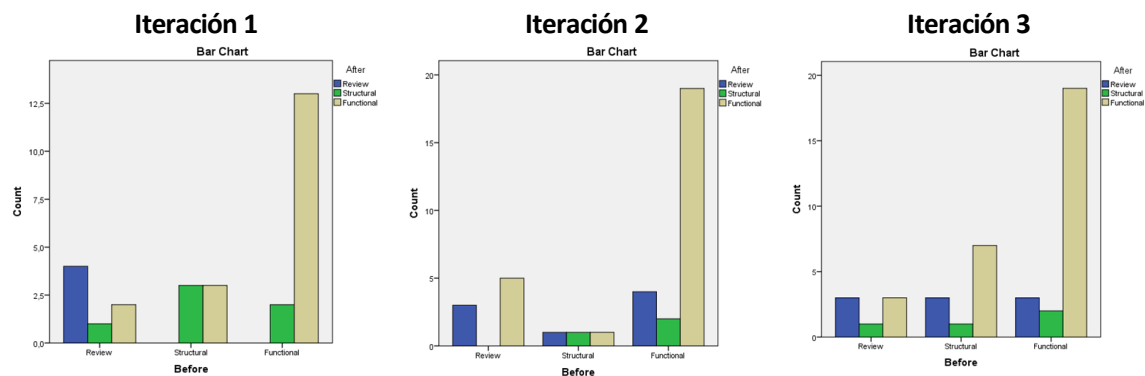


Figura 5.17. Cambios de opinión

Estos resultados se pueden contrastar en la Tabla 5.28, en la Tabla 5.29 y en la Tabla 5.30 para las Iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			After			Total
			Review	Structural	Functional	
Before	Review	Count	4	1	2	7
		% of Total	14,3%	3,6%	7,1%	25,0%
	Structural	Count	0	3	3	6
		% of Total	0,0%	10,7%	10,7%	21,4%
	Functional	Count	0	2	13	15
		% of Total	0,0%	7,1%	46,4%	53,6%
Total		Count	4	6	18	28
		% of Total	14,3%	21,4%	64,3%	100,0%

Tabla 5.28. Tabla de Contingencia de la iteración 1

			After			Total
			Review	Structural	Functional	
Before	Review	Count	3	0	5	8
		% of Total	8,3%	0,0%	13,9%	22,2%
	Structural	Count	1	1	1	3
		% of Total	2,8%	2,8%	2,8%	8,3%
	Functional	Count	4	2	19	25
		% of Total	11,1%	5,6%	52,8%	69,4%
Total		Count	8	3	25	36
		% of Total	22,2%	8,3%	69,4%	100,0%

Tabla 5.29. Tabla de Contingencia de la iteración 2

			After			Total
			Review	Structural	Functional	
Before	Review	Count	3	1	3	7
		% of Total	7,1%	2,4%	7,1%	16,7%
	Structural	Count	3	1	7	11
		% of Total	7,1%	2,4%	16,7%	26,2%
	Functional	Count	3	2	19	24
		% of Total	7,1%	4,8%	45,2%	57,1%
Total		Count	9	4	29	42
		% of Total	21,4%	9,5%	69,0%	100,0%

Tabla 5.30. Tabla de Contingencia de la iteración 3

5.1.3.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 71.4%, es decir la mayoría de los sujetos no cambian de opinión. Para la iteración 2 el porcentaje es del 63.9%, es decir la mayoría de los sujetos no cambian de opinión. Finalmente para la iteración 3 el porcentaje es del 54.7%, es decir la mayoría de los sujetos no cambian de opinión.

Sin embargo en la Tabla 5.31, se puede observar que para la iteración 1, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es

moderado ($>0,41 - 0,60$), se puede concluir que no existen cambios de opinión. Por otro lado, en las iteraciones 2 y 3, el acuerdo obtenido es debido al azar ($\text{sig}>0.05$), es decir, existen cambios de opinión.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,502	,144	3,722	,000
Iteración 2	,217	,153	1,634	,102
Iteración 3	,170	,112	1,589	,112

Tabla 5.31. Índice de Concordancia Kappa

5.1.3.2.2. Direccionalidad en el cambio de opinión

A continuación se muestran los resultados del Test de McNemar-Bowker. Según los resultados mostrados en la Tabla 5.32, para las tres iteraciones, hay simetría ($\text{sig}>0.05$) los cambios de una categoría a otra se producen en ambas direcciones con igual probabilidad, es decir, los cambios se producen de forma indistinta. En el caso de la iteración 1 se esperaba este resultado puesto que además existe acuerdo.

Iteración	Valor	gl	Sig. asintótica (bilateral)	N de casos válidos
Iteración 1	3,200	3	,362	28
Iteración 2	1,444	3	,695	36
Iteración 3	3,778	3	,286	42

Tabla 5.32. Pruebas de McNemar-Bowker

5.1.3.2.3. Homogeneidad Marginal

Finalmente se muestran los resultados del Test de Stuart-Maxwell. Según los resultados mostrados en la Tabla 5.33, para las tres iteraciones, la distribución de las tres técnicas es igual antes que después ($\text{sig}>0.05$), además se esperaba este resultado puesto que también existe simetría.

Iteración	χ^2	gl	Sig.
Iteración 1	3,176	2	,204
Iteración 2	0	2	1
Iteración 3	3,772	2	,152

Tabla 5.33. Prueba de Stuart-Maxwell

5.1.3.3. Discusión

Finalmente en la Tabla 5.34 se van a evaluar todos los test en conjunto para comprobar si realmente existen cambios de opinión y como son estos cambios.

Para la iteración 1, se puede observar que el acuerdo es moderado. Además se puede comprobar que hay simetría, es decir, cuando los sujetos cambian de opinión, el cambio es simétrico ($p_{ij}=p_{ji}$). Finalmente se comprueba que hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica es el mismo.

Para las iteraciones 2 y 3, se puede observar que no hay acuerdo. Además se puede comprobar que hay simetría, es decir, cuando los sujetos cambian de opinión, el cambio es simétrico ($p_{ij}=p_{ji}$). Finalmente se comprueba que hay homogeneidad marginal, es decir, el número de sujetos que responden antes una técnica determinada y después la misma técnica es el mismo.

	Acuerdo			Distribución Técnica				
	Kappa		McNemar-Bowker	Stuart-Maxwell	X ² (Antes)		X ² (Después)	
Iteración 1	,000	,502	,362	,204	,074	R: 25,00%	,001	R: 13,79%
						E: 21,43%		E: 20,69%
						F: 53,57%		F: 65,52%
Iteración 2	,102		,695	1	,000	R: 18,18%	,000	R: 23,68%
						E: 11,36%		E: 7,89%
						F: 70,45%		F: 68,42%
Iteración 3	,112		,286	0,152	,004	R: 16,67%	,000	R: 21,74%
						E: 26,19%		E: 8,7%
						F: 57,14%		F: 69,57%

Tabla 5.34. Tabla Resumen

5.2. Preferencias sobre los Programas

5.2.1. ¿Qué programa es más fácil?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa que consideran que es el más fácil. Solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

5.2.1.1. Distribución del Programa

Primero se va a analizar la distribución de los programas con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución del programa se realizará la prueba de distribución (X² de Pearson).

5.2.1.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.35 muestra los resultados de la categoría “Other” y la Tabla 5.36 muestra los resultados de la categoría “No answer”.

	Iteración 1	Iteración 2	Iteración 3
Respuestas	1 sujeto: más de un programa “Ntree/ Nametbl” 2 sujetos: “Igual”		1 sujeto: más de un programa “Nametbl/Ntree”
Total	3 sujetos	0 sujetos	1 sujeto

Tabla 5.35. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	5 sujetos
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 5.36. Sujetos en la categoría “No answer”

En la Figura 5.18, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1 se puede observar que la mayoría de los sujetos responden el

programa “Ntree”, además se puede observar que hay un gran porcentaje de sujetos en la categoría “No answer”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa “Nametbl”, además se puede observar que hay un gran porcentaje de sujetos en la categoría “No answer”. Finalmente, para la iteración 3 se puede observar que al igual que en la iteración 3 la mayoría de los sujetos responden el programa “Cmdline”, y hay muy pocos sujetos que responden otra cosa, es decir, pertenecen a la categoría “Other”.

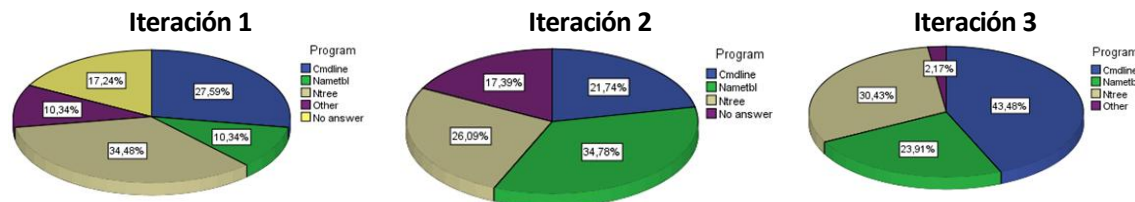


Figura 5.18. Después con categorías

5.2.1.1.2. Resultados sin categorías

En la Figura 5.19, para la iteración 1 se puede observar que la mayoría de los sujetos responden el programa “Ntree”, seguido de “Cmdline” y finalmente hay pocos sujetos que responden el programa “Nametbl”. Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa “Nametbl”, seguido de “Ntree” y finalmente hay pocos sujetos que responden el programa “Cmdline”. Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa “Cmdline”, seguido de “Ntree” y finalmente hay pocos sujetos que responden el programa “Nametbl”.



Figura 5.19. Después sin categorías

En la iteración 1 el programa que ocurre con mayor probabilidad es “Ntree”, en la iteración 2 es “Nametbl” y en la iteración 3 es “Cmdline”.

5.2.1.1.2.1. Prueba de proporciones

En la tabla 6.37, se puede observar que en las tres iteraciones las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una “preferencia definida” o “preferencia clara” hacia alguna de las técnicas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	21	3,714	2	,156
Iteración 2	38	1,474	2	,479
Iteración 3	45	2,800	2	,247

Tabla 5.37. Prueba de Chi-Cuadrado

En la Figura 5.20, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.38, para cada iteración.

	N observado			N esperado			Residual		
	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3	It. 1	It. 2	It. 3
Review	8	10	20	7,0	12,7	20	1,0	-2,7	5,0
Structural	3	16	11	7,0	12,7	11	-4,0	3,3	-4,0
Functiona	10	12	14	7,0	12,7	14	3,0	-,7	-1,0
Total	21	38	45						

Tabla 5.38. Frecuencias

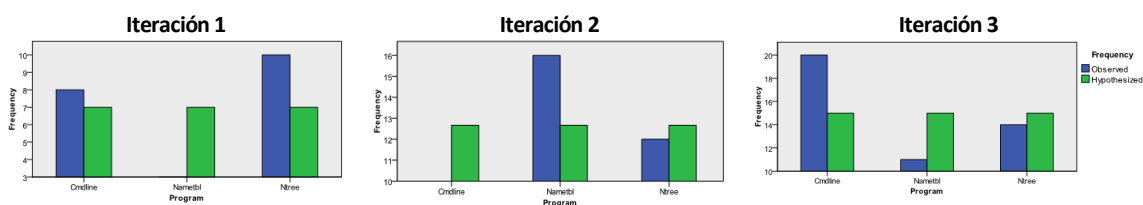


Figura 5.20. Chi-Cuadrado

5.2.2. ¿Qué programa tiene las faltas más difíciles?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa que consideran que tiene las faltas más difíciles. Solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

5.2.2.1. Distribución del Programa

Primero se va a analizar la distribución de los programas con las categorías “Other” y “No answer”. Luego se remplazarán los datos de dichas categorías por valores nulos y además de analizar la distribución del programa se realizará la prueba de distribución (χ^2 de Pearson).

5.2.2.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.39 muestra los resultados de la categoría “Other” y la Tabla 5.40 muestra los resultados de la categoría “No answer”.

	Iteración 1	Iteración 2	Iteración 3
Respuestas	11 sujetos: “Igual”		
Total	11 sujetos	0 sujetos	0 sujetos

Tabla 5.39. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 1	7 sujetos
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 5.40. Sujetos en la categoría "No answer"

En la Figura 5.21, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 1 se puede observar que la mayoría de los sujetos pertenecen a las categorías "Other" y "No answer"; el programa con mayor porcentaje de sujetos es "Ntree". Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa "Ntree", además se puede observar que hay un gran porcentaje de sujetos en la categoría "No answer". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa "Cmdline" y no se han creado las categorías "Other" ni "No answer".



Figura 5.21. Después con categorías

5.2.2.1.2. Resultados sin categorías

En la Figura 5.22, para la iteración 1 se puede observar que la mayoría de los sujetos responden el programa "Ntree", seguido de "Cmdline" y finalmente hay pocos sujetos que responden "Nametbl". Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa "Ntree", seguido de "Cmdline" y finalmente el resto de sujetos responden el programa "Nametbl". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa "Cmdline", seguido de "Nametbl" y finalmente de "Ntree".



Figura 5.22. Después sin categorías

En la iteración 1 y en la iteración 2 el programa que ocurre con mayor probabilidad es "Ntree", mientras que en la iteración 3 es "Cmdline".

5.2.2.1.2.1. Prueba de proporciones

En la tabla 6.41, se puede observar que en las tres iteraciones las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una "preferencia definida" o "preferencia clara" hacia alguna de las técnicas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 1	11	3,455	2	,178
Iteración 2	38	,368	2	,832
Iteración 3	46	,304	2	,859

Tabla 5.41. Prueba de Chi-Cuadrado

En la Figura 5.23, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.42, para cada iteración.

	N observado			N esperado			Residual		
	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3	lt. 1	lt. 2	lt. 3
Review	4	13	17	3,7	12,7	15,3	,3	,3	1,7
Structural	1	11	15	3,7	12,7	15,3	-2,7	-1,7	-,3
Functional	6	14	14	3,7	12,7	15,3	2,3	1,3	-1,3
Total	11	38	46						

Tabla 5.42. Frecuencias

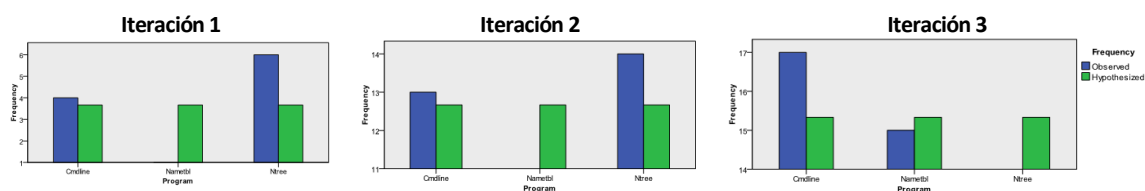


Figura 5.23. Chi-Cuadrado

5.2.3. ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa que consideran que lo han entendido mejor. Esta pregunta aparece únicamente en la segunda y tercera iteración, solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

5.2.3.1. Distribución del Programa

Primero se va a analizar la distribución de los programas con las categorías “Other” y “No answer”. Luego se remplazaran los datos de dichas categorías por valores nulos y además de analizar la distribución del programa se realizará la prueba de distribución (χ^2 de Pearson).

5.2.3.1.1. Resultados con Categorías “Other” y “No answer”

Para mejor apreciación y entendimiento, se muestran los resultados de las respuestas de antes en un gráfico circular. La Tabla 5.43 muestra los resultados de la categoría “Other” y la Tabla 5.44 muestra los resultados de la categoría “No answer”.

	Iteración 2	Iteración 3
Respuestas		1 sujeto: más de un programa “Nametbl/Ntree”
Total	0 sujetos	1 sujeto

Tabla 5.43. Respuestas para categoría “Other”

Iteración	Nº de sujetos
Iteración 2	8 sujetos
Iteración 3	0 sujetos

Tabla 5.44. Sujetos en la categoría "No answer"

En la Figura 5.24, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa "Nametbl", además se puede observar que hay un gran porcentaje de sujetos en la categoría "No answer". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa "Cmdline" y "Ntree" con el mismo porcentaje de sujetos, y hay muy pocos sujetos que responden otra cosa, es decir, pertenecen a la categoría "Other".

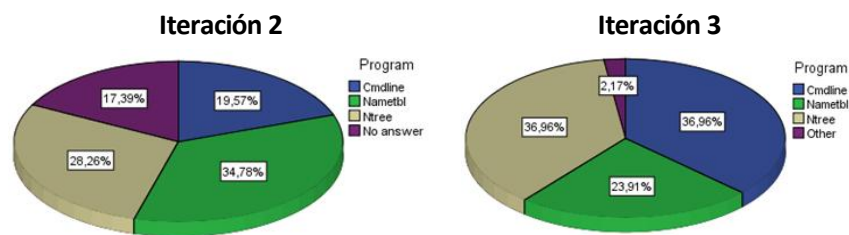


Figura 5.24. Después con categorías

5.2.3.1.2. Resultados sin categorías

En la Figura 5.25, para la iteración 1 se puede observar que la mayoría de los sujetos responden el programa "Ntree", seguido de "Cmdline" y finalmente hay pocos sujetos que responden el programa "Nametbl". Para la iteración 2 se puede observar que la mayoría de los sujetos responden el programa "Nametbl", seguido de "Ntree" y finalmente hay pocos sujetos que responden el programa "Cmdline". Finalmente, para la iteración 3 se puede observar que la mayoría de los sujetos responden el programa "Cmdline" y "Ntree" con el mismo porcentaje, seguido del programa "Nametbl".

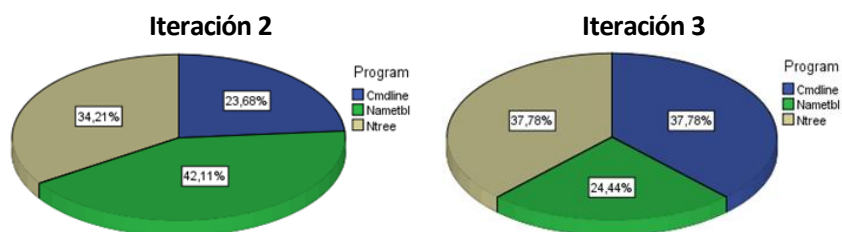


Figura 5.25. Después sin categorías

En la iteración 2 el programa que ocurre con mayor probabilidad es "Nametbl", mientras que en la iteración 3 es "Cmdline" y "Ntree" pues ambos tienen el mismo porcentaje.

5.2.3.1.2.1. Prueba de proporciones

En la tabla 6.45, se puede observar que en las dos iteraciones las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una "preferencia definida" o "preferencia clara" hacia alguna de las técnicas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 2	38	1,947	2	,378
Iteración 3	45	1,600	2	,449

Tabla 5.45. Prueba de Chi-Cuadrado

En la Figura 5.26, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.46, para cada iteración.

	N observado		N esperado		Residual	
	It. 2	It. 3	It. 2	It. 3	It. 2	It. 3
Review	9	17	12,7	15,0	-3,7	2,0
Structural	16	11	12,7	15,0	3,3	-4,0
Functional	13	17	12,7	15,0	,3	2,0
Total	38	45				

Tabla 5.46. Frecuencias

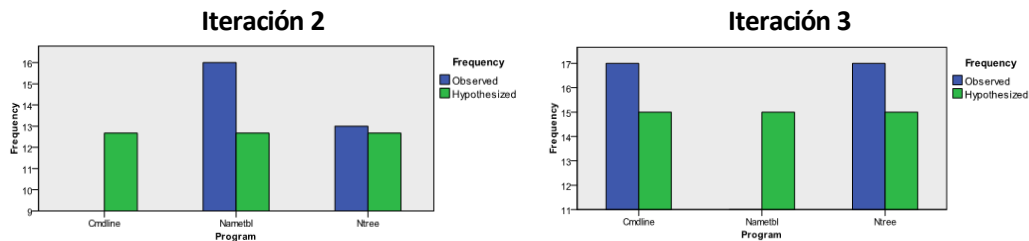


Figura 5.26. Chi-Cuadrado

5.2.4. ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la opinión de los sujetos con respecto al programa que consideran que tiene las faltas más difíciles. Esta pregunta solo aparece en la tercera iteración, solo hay respuestas de después de aplicar las técnicas, dado que la pregunta solo tiene sentido después de aplicarlas.

5.2.4.1. Distribución del Programa

5.2.4.1.1. Resultados sin categorías

Dado que no aparecen respuestas raras y además todos los sujetos responden, no ha sido necesario crear las categorías de “Other” ni “No answer”.

En la Figura 5.27, se pueden observar el porcentaje de sujetos que pertenecen a la misma categoría. Para la iteración 3 la mayoría de los sujetos responden el programa “Cmdline”, seguido del programa “Nametbl” y finalmente el resto de sujetos responden el programa “Ntree”.



Figura 5.27. Después sin categorías

5.2.4.1.1.1. Prueba de proporciones

En la Tabla 5.47, se puede observar que en la iteración 3 las tres técnicas ocurren con la misma probabilidad ($\text{sig} > 0.05$), lo que significa que, no hay una “preferencia definida” o “preferencia clara” hacia alguna de las técnicas.

Iteración	N Total	Prueba Estadística	gl	Sig. asintótica (bilateral)
Iteración 3	46	4,217	2	,121

Tabla 5.47. Prueba de Chi-Cuadrado

En la Figura 5.28, se puede observar el valor observado en azul y el esperado en verde de las frecuencias de las técnicas mostradas en la Tabla 5.48, para cada iteración.

	N observado	N esperado	Residual
Review	20	15,3	4,7
Structural	17	15,3	1,7
Functional	9	15,3	-6,3
Total	46		

Tabla 5.48. Frecuencias

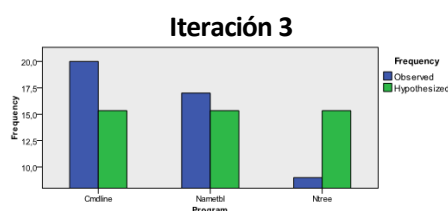


Figura 5.28. Chi-Cuadrado

5.3. Consistencia en las Contestaciones sobre las Técnicas

En este capítulo se va a analizar la relación existente entre las preguntas de preferencias entre sí. Para ello se seguirá el método explicado en el capítulo 2. Se va a comparar cada variable de la fila de la izquierda con las variables que aparecen en la derecha de la Tabla 5.49.

Pregunta 1	Pregunta 2
Técnica preferida	Técnica más fácil de entender
	Técnica más fácil de aplicar
Técnica más fácil de entender	Técnica más fácil de aplicar
Programa más fácil	Programa con faltas más difíciles
	Programa más fácil de entender
	Programa con faltas más fáciles
Programa con faltas más difíciles	Programa más fácil de entender
Programa más fácil de entender	Programa con faltas más fáciles
	Programa con faltas más fáciles

Tabla 5.49. Relación entre las preguntas de Preferencias

5.3.1. ¿Qué técnica te gusta más? / ¿Qué Técnica te ha parecido más fácil de entender?

En esta sección se va a evaluar la relación existente entre la técnica que más les gusta a los sujetos y la técnica que creen que han entendido mejor. Esta pregunta aparece en las tres iteraciones. Dado que la segunda pregunta solo se realizó después del estudio empírico, se va a analizar la combinación: Después-Después.

5.3.1.1. Estudio del Acuerdo Después-Después

En la Figura 5.29, para la iteración 1, se observa que ninguna persona considera que la técnica que más les gusta es “Revisión”. De cuatro personas que dicen que la técnica que más les gusta es “Estructural”, uno también la considera la técnica mejor entendida y para tres la mejor entendida es “Revisión”. Finalmente de catorce personas que dicen que la técnica que más les gusta es “Funcional” tres también la consideran la técnica mejor entendida, de los once restantes: seis consideran que la mejor entendida es “Revisión” y los otros cinco “Estructural”.

Para la iteración 2, de ocho personas que dicen que su técnica preferida es “Revisión”, cuatro también la consideran la técnica mejor entendida, de los cuatro restantes: para tres la técnica mejor entendida es “Estructural” y para uno es “Funcional”. De tres personas que dicen que su técnica preferida es “Estructural” ninguna la considera la técnica mejor entendida, los tres sujetos consideran que la mejor entendida es “Funcional”. Finalmente de veinticinco personas que dicen que su técnica preferida es “Funcional” catorce también la consideran la técnica mejor entendida, de los once restantes: para ocho la mejor entendida es “Revisión” y para cinco “Estructural”.

Para la iteración 3, de trece personas que dicen que la técnica que más les gusta es “Revisión”, nueve también la consideran la técnica mejor entendida, de los cuatro restantes: para uno la técnica mejor entendida es “Estructural” y para tres es “Funcional”. De tres personas que dicen que la técnica que más les gusta es “Estructural” uno también la considera la técnica mejor entendida, de los dos restantes: uno considera que la técnica mejor entendida es “Revisión” y otro “Funcional”. Finalmente de veinticuatro personas que dicen que la técnica que más les gusta es “Funcional” nueve también la consideraban la técnica mejor entendida, de los quince restantes: para nueve la técnica mejor entendida es “Revisión” y para los otros seis “Estructural”.

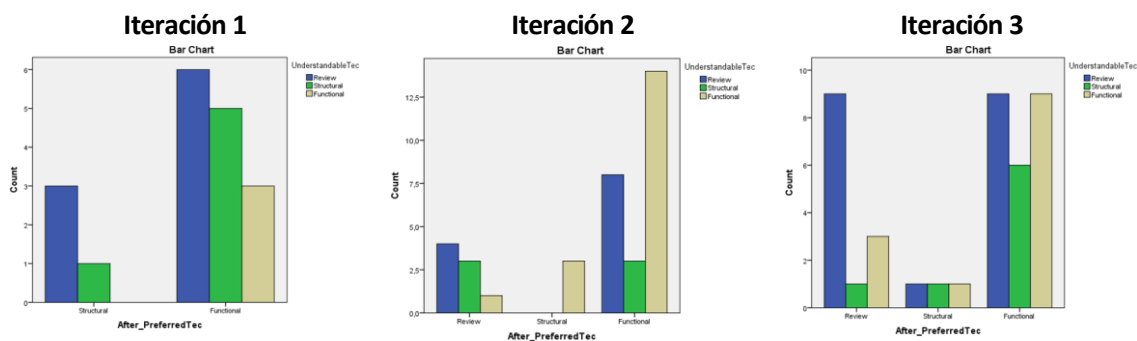


Figura 5.29. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.50, en la Tabla 5.51 y en la Tabla 5.52 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PT2			Total
			Review	Structural	Functiona	
PT1(After)	Structural	Count	3	1	0	4
		% of Total	16,7%	5,6%	0,0%	22,2%
	Functional	Count	6	5	3	14
		% of Total	33,3%	27,8%	16,7%	77,8%
Total		Count	9	6	3	18
		% of Total	50,0%	33,3%	16,7%	100,0%

Tabla 5.50. Tabla de Contingencia: de la iteración 1

			PT2			Total
			Review	Structural	Functiona	
PT1(After)	Review	Count	4	3	1	8
		% of Total	11,1%	8,3%	2,8%	22,2%
	Structural	Count	0	0	3	3
		% of Total	0,0%	0,0%	8,3%	8,3%
	Functional	Count	8	3	14	25
		% of Total	22,2%	8,3%	38,9%	69,4%
Total		Count	12	6	18	36
		% of Total	33,3%	16,7%	50,0%	100,0%

Tabla 5.51. Tabla de Contingencia: de la iteración 2

			PT2			Total
			Review	Structural	Functiona	
PT1(After)	Review	Count	9	1	3	13
		% of Total	22,5%	2,5%	7,5%	32,5%
	Structural	Count	1	1	1	3
		% of Total	2,5%	2,5%	2,5%	7,5%
	Functional	Count	9	6	9	24
		% of Total	22,5%	15,0%	22,5%	60,0%
Total		Count	19	8	13	40
		% of Total	47,5%	20,0%	32,5%	100,0%

Tabla 5.52. Tabla de Contingencia: de la iteración 3

5.3.1.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 22.3%, es decir muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es del 50%, es decir la mitad de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 47.5%, es decir menos de la mitad de los sujetos están de acuerdo.

En la Tabla 5.53, para las tres iteraciones se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,023	,065	,275	,783
Iteración 2	,115	,115	,957	,339
Iteración 3	,174	,111	1,609	,108

Tabla 5.53. Índice de Concordancia Kappa

Debido a que para las tres iteraciones, no hay acuerdo, se puede concluir que los sujetos no identifican la técnica que les gusta más con la técnica más fácil de entender.

5.3.2. ¿Qué técnica te gusta más? / ¿Qué técnica te ha parecido más fácil de aplicar?

En esta sección se va a evaluar la relación existente entre la técnica que más les gusta a los sujetos y la técnica que ellos consideran que es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones. Dado que ambas preguntas se realizaron antes y después del estudio empírico, se va a analizar las combinaciones: Antes-Antes y Después-Después.

5.3.2.1. Estudio del Acuerdo Antes-Antes

En la Figura 5.30, para la iteración 1, se observa que de dos personas que dicen que su técnica preferida es “Revisión”, ninguna la considera la técnica más fácil de aplicar, por lo tanto los tres sujetos se distribuyen así: para uno la técnica más fácil de aplicar es “Estructural” y para otro “Funcional”. De catorce personas que dicen que su técnica preferida es “Estructural” tres también la consideran la técnica más fácil de aplicar, de los once restantes: cinco consideran que la técnica más fácil de aplicar es “Revisión” y los otros seis “Funcional”. Finalmente, de diez personas que su técnica preferida es “Funcional” ocho también la consideran la técnica más fácil de aplicar y los otros dos restantes: para uno la más fácil de aplicar es “Revisión” y para otro “Estructural”.

Para la iteración 2, de ocho personas que dicen que su técnica preferida es “Revisión”, todos también la consideran la más fácil de aplicar. De cinco personas que dicen que su técnica preferida es “Estructural”, todos también la consideran la más fácil de aplicar. Finalmente, de treinta y un personas que su técnica preferida es “Funcional”, todos además la consideran la más fácil de aplicar.

Para la iteración 3, de diez personas que dicen que su técnica preferida es “Revisión”, tres también la consideran la más fácil de aplicar, de los siete restantes: para cuatro la técnica más fácil de aplicar es “Estructural” y para tres “Funcional”. De trece personas que dicen que su técnica es “Estructural” cinco también la consideran la más fácil de aplicar, de los ocho restantes: dos consideran que la técnica más fácil de aplicar es “Revisión” y los otros seis “Funcional”. Finalmente, de diecinueve personas que su técnica preferida es “Funcional” quince también la consideran la más fácil de aplicar y de los otros cuatro restantes: para dos la técnica más fácil de aplicar es “Revisión” y para dos “Estructural”.

Estos resultados se pueden contrastar en la Tabla 5.54, Tabla 5.55 y Tabla 5.56 para las iteraciones 1, 2 y 3, respectivamente. Donde se indica el porcentaje de sujetos de cada técnica.

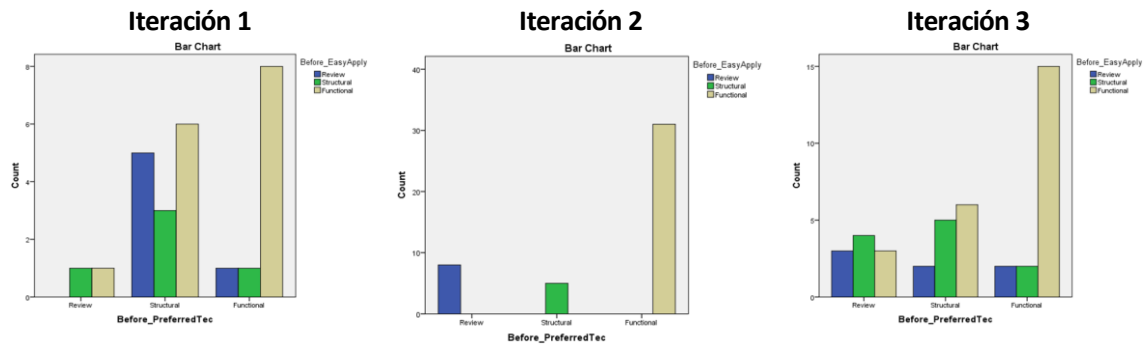


Figura 5.30. Acuerdo entre las respuestas

			PT3(Before)			Total
			Review	Structural	Functional	
PT1(Before)	Review	Count	0	1	1	2
		% of Total	0,0%	3,8%	3,8%	7,7%
	Structural	Count	5	3	6	14
		% of Total	19,2%	11,5%	23,1%	53,8%
	Functional	Count	1	1	8	10
		% of Total	3,8%	3,8%	30,8%	38,5%
Total		Count	6	5	15	26
		% of Total	23,1%	19,2%	57,7%	100,0%

Tabla 5.54. Tabla de Contingencia de la iteración 1

			PT3(Before)			Total
			Review	Structural	Functional	
PT1(Before)	Review	Count	8	0	0	8
		% of Total	18,1%	0,0%	0,0%	18,1%
	Structural	Count	0	5	0	5
		% of Total	0,0%	11,4%	0,0%	11,4%
	Functional	Count	0	0	31	31
		% of Total	0,0%	0,0%	70,5%	70,5%
Total		Count	8	5	31	44
		% of Total	18,2%	11,4%	70,5%	100,0%

Tabla 5.55. Tabla de Contingencia de la iteración 2

			PT3(Before)			Total
			Review	Structural	Functional	
PT1(Before)	Review	Count	3	4	3	10
		% of Total	7,1%	9,5%	7,1%	23,8%
	Structural	Count	2	5	6	13
		% of Total	4,8%	11,9%	14,3%	31,0%
	Functional	Count	2	2	15	19
		% of Total	4,8%	4,8%	35,7%	45,2%
Total		Count	7	11	24	42
		% of Total	16,7%	26,2%	57,1%	100,0%

Tabla 5.56. Tabla de Contingencia de la iteración 3

5.3.2.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 42.3%, es decir casi la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es del 100%, es decir, todos los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 54.7%, es decir la mayoría de los sujetos están de acuerdo.

En la Tabla 5.57, para la iteración 1 se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo. Para la iteración 2 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es perfecto (0,81 - 1), se puede concluir que existe acuerdo. Finalmente, para la iteración 3 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es discreto (0,21 - 0,40), se puede concluir que existe poco acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,122	,111	,998	,318
Iteración 2	1,000	,000	8,688	,000
Iteración 3	,271	,113	2,464	,014

Tabla 5.57. Índice de Concordancia Kappa

Debido a que para las iteración 1 no hay acuerdo, sin embargo en las otras dos iteraciones si lo hay y en una de las mismas el acuerdo es absoluto, se puede concluir que en cierta manera los sujetos identifican la técnica que les gusta más con la técnica más fácil de aplicar. La relación entre las preguntas es intermedia.

5.3.2.2. Estudio del Acuerdo Después - Después

En la Figura 5.31, para la iteración 1, se observa que de una persona que dijo que la técnica que más les gusta es "Revisión", ninguna la considera la técnica más fácil de aplicar, por lo tanto el sujeto considera que la técnica más fácil de aplicar es "Funcional". De siete personas que dicen que la técnica que más les gusta es "Estructural" dos también la consideran la técnica más fácil de aplicar, de los cinco restantes: dos consideran que la técnica más fácil de aplicar es "Revisión" y los otros tres "Funcional". Finalmente, de dieciocho personas que dicen que la técnica que más les gusta es "Funcional" trece también la consideran la técnica más fácil de aplicar y de los otros cinco restantes: para uno la técnica más fácil de aplicar es "Revisión" y para los otros cuatro es "Estructural".

Para la iteración 2, de nueve personas que dicen que la técnica que más les gusta es "Revisión", todos también la consideran la técnica más fácil de aplicar. De tres personas que dicen que la técnica que más les gusta es "Estructural", todos también la consideran la técnica más fácil de aplicar. Finalmente, de veintiséis personas que la técnica que más les gusta es "Funcional", todos además la consideran la técnica más fácil de aplicar.

Para la iteración 3, de catorce personas que dicen que la técnica que más les gusta es "Revisión", seis también la consideran la técnica más fácil de aplicar, de los ocho restantes: para uno la técnica más fácil de aplicar "Estructural" y para siete "Funcional". De cuatro personas que dicen que la técnica que más les gusta es "Estructural" dos también la consideran la técnica más fácil de

aplicar y los dos restantes consideran que la técnica más fácil de aplicar es “Funcional”. Finalmente, de veintisiete personas que la técnica que más les gusta es “Funcional” veintidós también la consideran la técnica más fácil de aplicar, de los cinco restantes: para cuatro la técnica más fácil de aplicar es “Revisión” y para uno “Estructural”.

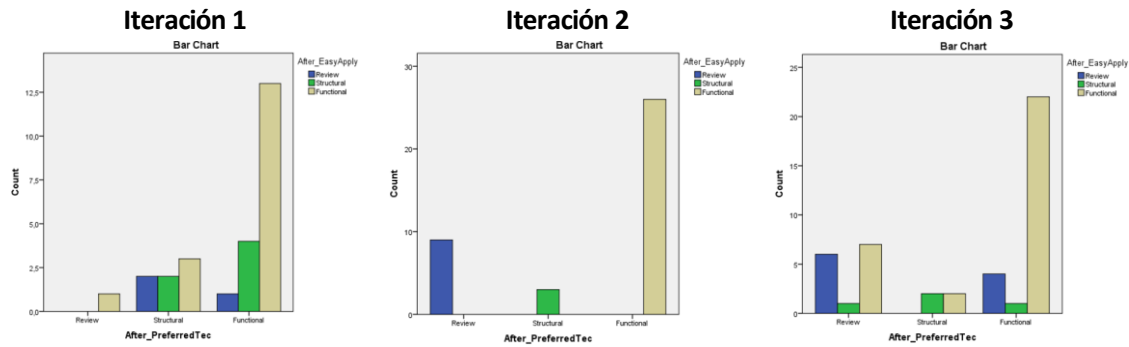


Figura 5.31. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.58, en la Tabla 5.59 y en la Tabla 5.60 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PT3(After)			Total
			Review	Structural	Functional	
PT1(After)	Review	Count	0	0	1	1
		% of Total	0,0%	0,0%	3,8%	3,8%
	Structural	Count	2	2	3	7
		% of Total	7,7%	7,7%	11,5%	26,9%
	Functional	Count	1	4	13	18
		% of Total	3,8%	15,4%	50,0%	69,2%
Total		Count	3	6	17	26
		% of Total	11,5%	23,1%	65,4%	100,0%

Tabla 5.58. Tabla de Contingencia de la iteración 1

			PT3(After)			Total
			Review	Structural	Functional	
PT1(After)	Review	Count	9	0	0	9
		% of Total	23,7%	0,0%	0,0%	23,7%
	Structural	Count	0	3	0	3
		% of Total	0,0%	7,9%	0,0%	7,9%
	Functional	Count	0	0	26	26
		% of Total	0,0%	0,0%	68,4%	68,4%
Total		Count	9	3	26	38
		% of Total	23,7%	7,9%	68,4%	100,0%

Tabla 5.59. Tabla de Contingencia de la iteración 2

			PT3(After)			Total
			Review	Structural	Functional	
PT1(After)	Review	Count	6	1	7	14
		% of Total	13,3%	2,2%	15,6%	31,1%
	Structural	Count	0	2	2	4
		% of Total	0,0%	4,4%	4,4%	8,9%
	Functional	Count	4	1	22	27
		% of Total	8,9%	2,2%	48,9%	60,0%
Total		Count	10	4	31	45
		% of Total	22,2%	8,9%	68,9%	100,0%

Tabla 5.60. Tabla de Contingencia de la iteración 3

5.3.2.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 57.7%, es decir más de la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es del 100%, es decir, todos los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 66.6%, es decir la mayoría de los sujetos están de acuerdo.

En la Tabla 5.61, para la iteración 1 se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo. Para la iteración 2 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es perfecto (0,81 - 1), se puede concluir que existe acuerdo. Finalmente, para la iteración 3 el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de kappa muestra que el acuerdo es discreto (0,21 - 0,40), se puede concluir que existe poco acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. Aproximada
Iteración 1	,120	,154	,765	,444
Iteración 2	1,000	,000	7,635	,000
Iteración 3	,346	,131	2,958	,003

Tabla 5.61. Índice de Concordancia Kappa

Debido a que para las iteración 1 no hay acuerdo, sin embargo en las otras dos iteraciones si lo hay y en una de las mismas el acuerdo es absoluto, se puede concluir que en cierta manera los sujetos identifican la técnica que les gusta más con la técnica más fácil de aplicar. La relación entre las preguntas es intermedia.

5.3.3. ¿Qué técnica te ha parecido más fácil de entender? / ¿Qué técnica te ha parecido más fácil de aplicar?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que han entendido mejor y la técnica que ellos consideran que es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones. Dado que la primera pregunta solo se realiza después del estudio empírico, se va a analizar las combinaciones: Después-Después.

5.3.3.1. Estudio del Acuerdo Después - Después

En la Figura 5.32, para la iteración 1, se observa que de nueve personas que dicen que la técnica más fácil de entender es “Revisión”, dos también la consideran la técnica más fácil de aplicar, de los siete restantes: para dos la técnica más fácil de aplicar “Estructural” y para cinco “Funcional”. De seis personas que dicen que la técnica más fácil de entender es “Estructural” uno también la considera la técnica más fácil de aplicar, de los cinco restantes: uno considera que la técnica más fácil de aplicar es “Revisión” y los otros cuatro “Funcional”. Finalmente, de tres personas que la técnica más fácil de entender es “Funcional” dos también la consideran la técnica más fácil de aplicar y la persona restante considera que es “Estructural”.

Para la iteración 2, de doce personas que dicen que la técnica que más fácil de entender es “Revisión”, cuatro también la consideran la técnica más fácil de aplicar y los ocho restantes consideran que la técnica más fácil de aplicar es “Funcional”. De seis personas que dicen que la técnica que más fácil de entender es “Estructural”, ninguna la consideran la técnica más fácil de aplicar, por lo tanto los seis sujetos se distribuyen de la siguiente manera: tres consideran que la técnica más fácil de aplicar es “Revisión” y tres “Funcional”. Finalmente, de dieciocho personas que la técnica que más fácil de entender es “Funcional”, catorce también la consideran la técnica más fácil de aplicar, de los otros cuatro restantes: para uno la técnica más fácil de aplicar es “Revisión” y para tres “Estructural”.

Para la iteración 3, de diecinueve personas que dicen que la técnica que más fácil de entender es “Revisión”, seis también la consideran la técnica más fácil de aplicar, de los trece restantes: para dos la técnica más fácil de aplicar “Estructural” y para once “Funcional”. De nueve personas que dicen que la técnica que más fácil de entender es “Estructural” una también la consideran la técnica más fácil de aplicar, de los ocho restantes: uno consideran que la técnica más fácil de aplicar es “Revisión” y los otros siete “Funcional”. Finalmente, de trece personas que dicen que la técnica que más fácil de entender es “Funcional” diez también la consideran la técnica más fácil de aplicar, de los tres restantes: para dos la técnica más fácil de aplicar es “Revisión” y para el otro sujeto es “Estructural”.

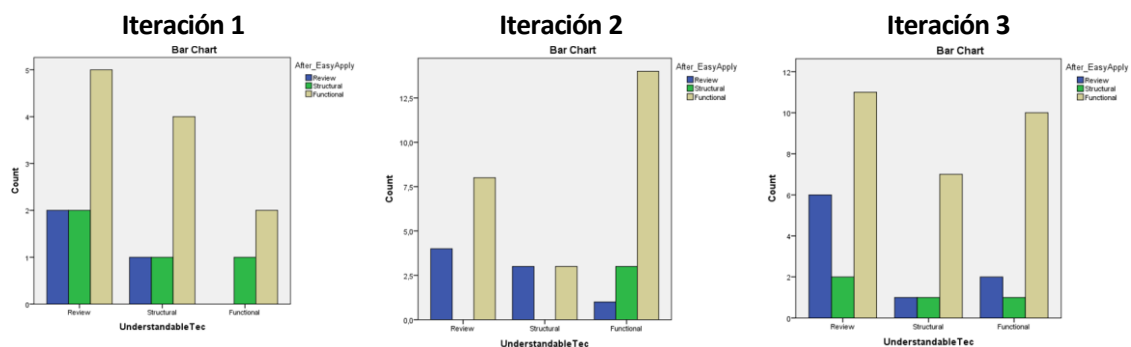


Figura 5.32. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.62, en la Tabla 5.63 y en la Tabla 5.64 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PT3(After)			Total
			Review	Structural	Functional	
PT2	Review	Count	2	2	5	9
		% of Total	11,1%	11,1%	27,8%	50,0%
	Structural	Count	1	1	4	6
		% of Total	5,6%	5,6%	22,2%	33,3%
	Functional	Count	0	1	2	3
		% of Total	0,0%	5,6%	11,1%	16,7%
Total		Count	3	4	11	18
		% of Total	16,7%	22,2%	61,1%	100,0%

Tabla 5.62. Tabla de Contingencia de la iteración 1

			PT3(After)			Total
			Review	Structural	Functional	
PT2	Review	Count	4	0	8	12
		% of Total	11,1%	0,0%	22,2%	33,3%
	Structural	Count	3	0	3	6
		% of Total	8,3%	0,0%	8,3%	16,7%
	Functional	Count	1	3	14	18
		% of Total	2,8%	8,3%	38,9%	50,0%
Total		Count	8	3	25	36
		% of Total	22,2%	8,3%	69,4%	100,0%

Tabla 5.63. Tabla de Contingencia de la iteración 2

			PT3(After)			Total
			Review	Structural	Functional	
PT2	Review	Count	6	2	11	19
		% of Total	14,6%	4,9%	26,8%	46,3%
	Structural	Count	1	1	7	9
		% of Total	2,4%	2,4%	17,1%	22,0%
	Functional	Count	2	1	10	13
		% of Total	4,9%	2,4%	24,4%	31,7%
Total		Count	9	4	28	41
		% of Total	22,0%	9,8%	68,3%	100,0%

Tabla 5.64. Tabla de Contingencia de la iteración 3

5.3.3.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 27.8%, es decir muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es del 50%, es decir, la mitad de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 41.4%, es decir pocos sujetos están de acuerdo.

En la Tabla 6.65, para las tres iteraciones se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo.

Debido a que en ninguna de las tres iteraciones hay acuerdo, se puede concluir que los sujetos no identifican la técnica que les ha parecido más fácil de entender con la técnica les ha parecido más fácil de aplicar.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. Aproximada
Iteración 1	,025	,120	,202	,840
Iteración 2	,115	,115	,957	,339
Iteración 3	,114	,097	1,191	,233

Tabla 5.65. Índice de Concordancia Kappa

5.4. Consistencias en las Contestaciones sobre los Programas

5.4.1. ¿Qué programa es más fácil? / ¿Qué programa tiene las faltas más difíciles?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que es el más fácil y el programa que tiene las faltas más difíciles. Esta pregunta aparece en las tres iteraciones. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.1.1. Estudio del Acuerdo Después - Después

En la Figura 5.33, para la iteración 1, se observa que de cinco personas que dicen que el programa más fácil es "Cmdline", uno también lo considera el programa con faltas más difíciles, de los cuatro restantes: uno considera que el programa con faltas más difíciles es "Nmetbl" y tres "Ntree". De dos personas que dicen que el programa más fácil es "Nmetbl", ninguna lo considera el programa con faltas más difíciles, por lo tanto los dos sujetos se distribuyen así: uno considera que el programa con faltas más difíciles es "Cmdline" y otro "Ntree". Finalmente, de cuatro personas que dicen que el programa más fácil es "Ntree", dos también lo consideran el programa con faltas más difíciles y los otros dos consideran que el programa con faltas más difíciles es "Cmdline".

Para la iteración 2, de diez personas que dicen que el programa más fácil es "Cmdline", ninguna lo considera el programa con faltas más difíciles, por lo tanto los diez sujetos se distribuyen así: dos consideran que el programa con faltas más difíciles es "Nmetbl" y ocho "Ntree". De dieciséis personas que dicen que el programa más fácil es "Nmetbl", cuatro también lo consideran el programa con faltas más difíciles, de los doce restantes: seis consideran que el programa con faltas más difíciles es "Cmdline" y otros seis "Ntree". Finalmente, de doce personas que dicen que el programa más fácil es "Ntree", ninguno lo considera el programa con faltas más difíciles, por lo tanto los doce sujetos se distribuyen así: siete consideran que el programa con faltas más difíciles es "Cmdline" y los otros cinco "Nmetbl".

Para la iteración 3, de veinte personas que dicen que el programa más fácil es "Cmdline", dos también lo consideran el programa con faltas más difíciles, de los dieciocho restantes: nueve consideran que el programa con faltas más difíciles es "Nmetbl" y otros nueve "Ntree". De once

personas que dicen que el programa más fácil es “Nametbl”, una también lo considera el programa con faltas más difíciles, de los diez restantes: seis consideran que el programa con faltas más difíciles es “Cmdline” y otros cuatro “Ntree”. Finalmente, de catorce personas que dicen que el programa más fácil es “Ntree”, una también lo considera el programa con faltas más difíciles, de los trece restantes: ocho consideran que el programa con faltas más difíciles es “Cmdline” y los otros cinco responden “Nametbl”.

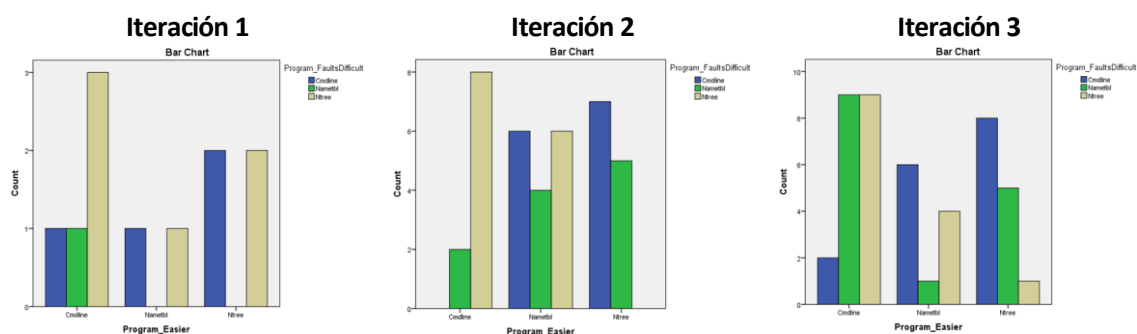


Figura 5.33. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.66, en la Tabla 5.67 y en la Tabla 5.68 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a qué porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP2			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	1	1	3	5
		% of Total	9,1%	9,1%	27,3%	45,5%
	Nametbl	Count	1	0	1	2
		% of Total	9,1%	0,0%	9,1%	18,2%
	Ntree	Count	2	0	2	4
		% of Total	18,2%	0,0%	18,2%	36,4%
Total		Count	4	1	6	11
		% of Total	36,4%	9,1%	54,5%	100,0%

Tabla 5.66. Tabla de Contingencia de la iteración 1

			PP2			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	0	2	8	10
		% of Total	0,0%	5,3%	21,1%	26,3%
	Nametbl	Count	6	4	6	16
		% of Total	15,8%	10,5%	15,8%	42,1%
	Ntree	Count	7	5	0	12
		% of Total	18,4%	13,2%	0,0%	31,6%
Total		Count	13	11	14	38
		% of Total	34,2%	28,9%	36,8%	100,0%

Tabla 5.67. Tabla de Contingencia de la iteración 2

			PP2			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	2	9	9	20
		% of Total	4,4%	20,0%	20,0%	44,4%
	Nametbl	Count	6	1	4	11
		% of Total	13,3%	2,2%	8,9%	24,4%
	Ntree	Count	8	5	1	14
		% of Total	17,8%	11,1%	2,2%	31,1%
Total		Count	16	15	14	45
		% of Total	35,6%	33,3%	31,1%	100,0%

Tabla 5.68. Tabla de Contingencia de la iteración 3

5.4.1.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 27.3%, es decir muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es del 10.5%, es decir, muy pocos sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 8.8%, es decir pocos sujetos están de acuerdo. Se esperaban estos porcentajes tan bajos, puesto que las preguntas pueden considerarse contradictorias.

En la Tabla 5.69, para la iteración 1, se puede observar el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), es decir, no existe acuerdo. Tanto para la iteración 2 como para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa es negativo se interpretará el valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	-,173	,199	-,788	,431
Iteración 2	-,332	,072	-2,962	,003
Iteración 3	-,373	,066	-3,574	,000

Tabla 5.69. Índice de Concordancia Kappa

Debido a que en la iteración 1 no hay acuerdo y en las iteraciones 2 y 3 hay un desacuerdo discreto, se puede concluir que, como era de esperar, los sujetos no identifican el programa más fácil con el programa que tiene las faltas más difíciles.

5.4.2. ¿Qué programa es más fácil? / ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que es el más fácil y el programa que mejor han entendido. La segunda pregunta solo aparece en la segunda y tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.2.1. Estudio del Acuerdo Después – Después

En la Figura 5.34, para la iteración 2, de diez personas que dicen que el programa más fácil es “Cmdline”, ocho también lo consideran el mejor entendido, de los dos restantes: uno considera que el mejor entendido es “Nametbl” y otro “Ntree”. De dieciséis personas que dicen que el programa más fácil es “Nametbl”, trece también lo consideran el programa mejor entendido y para los otros tres el mejor entendido es “Ntree”. Finalmente, de doce personas que dicen que el programa más fácil es “Ntree”, nueve también lo consideran el programa mejor entendido, de los tres restantes: uno considera que el programa mejor entendido es “Cmdline” y los otros dos “Nametbl”.

Para la iteración 3, de veinte personas que dicen que el programa más fácil es “Cmdline”, quince también lo consideran el programa mejor entendido, de los cinco restantes: dos consideran que el programa mejor entendido es “Nametbl” y otros tres “Ntree”. De once personas que dicen que el programa más fácil es “Nametbl”, siete también lo considera el programa mejor entendido, de los cuatro restantes: uno considera que el programa mejor entendido es “Cmdline” y otros tres “Ntree”. Finalmente, de catorce personas que dicen que el programa más fácil es “Ntree”, once también lo consideran el programa mejor entendido, de los tres restantes: uno consideran que el programa mejor entendido es “Cmdline” y los otros dos “Nametbl”.

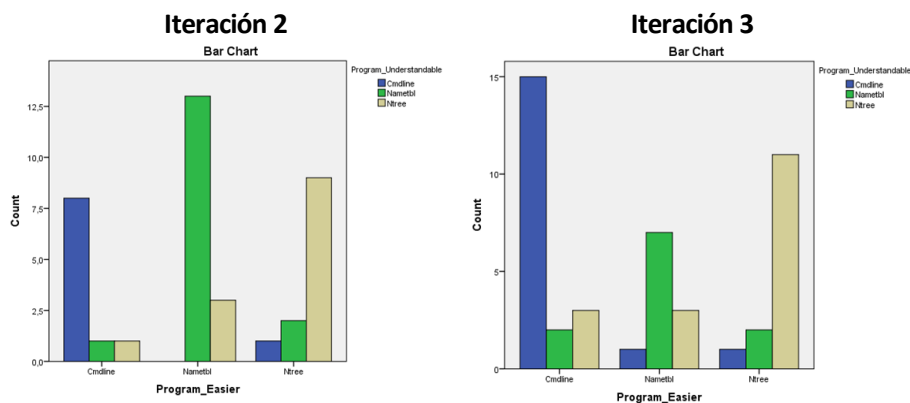


Figura 5.34. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.70 y en la Tabla 5.71 para las Iteraciones 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP3			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	8	1	1	10
		% of Total	21,1%	2,6%	2,6%	26,3%
	Nametbl	Count	0	13	3	16
		% of Total	0,0%	34,2%	7,9%	42,1%
	Ntree	Count	1	2	9	12
		% of Total	2,6%	5,3%	23,7%	31,6%
Total		Count	9	16	13	38
		% of Total	23,7%	42,1%	34,2%	100,0%

Tabla 5.70. Tabla de Contingencia de la iteración 2

			PP3			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	15	2	3	20
		% of Total	33,3%	4,4%	6,7%	44,4%
	Nametbl	Count	1	7	3	11
		% of Total	2,2%	15,6%	6,7%	24,4%
	Ntree	Count	1	2	11	14
		% of Total	2,2%	4,4%	24,4%	31,1%
Total		Count	17	11	17	45
		% of Total	37,8%	24,4%	37,8%	100,0%

Tabla 5.71. Tabla de Contingencia de la iteración 3

5.4.2.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 2 es del 79%, es decir la mayoría de sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 73.3%, es decir la mayoría de sujetos están de acuerdo.

En la Tabla 6.72, para la iteración 2, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el resultado de Kappa muestra que el acuerdo es sustancial (0,61 - 0,80), se puede concluir que existe acuerdo. Finalmente, para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el resultado de Kappa muestra que el acuerdo es moderado ($> 0,41$ - 0,60), se puede concluir que existe acuerdo

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 2	,677	,102	5,847	,000
Iteración 3	,593	,099	5,601	,000

Tabla 5.72. Índice de Concordancia Kappa

Debido a que en la iteración 2 hay un acuerdo sustancial y en la iteración 3 hay un acuerdo moderado, se puede concluir que, los sujetos identifican el programa más fácil con el programa que han entendido mejor. La relación entre las preguntas es intermedia.

5.4.3. ¿Qué programa tiene las faltas más difíciles? / ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que tiene las faltas más difíciles y el programa que mejor han entendido. La primera pregunta solo aparece en la segunda y tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.3.1. Estudio del Acuerdo Después-Después

En la Figura 5.35, para la iteración 2, de trece personas que dicen que el programa con faltas más difíciles es "Cmdline", ninguna lo consideran el programa mejor entendido, por lo tanto los

trece sujetos se distribuyen así: seis consideran que el mejor entendido es “Nametbl” y los otros siete “Ntree”. De once personas que dicen que el programa con faltas más difíciles es “Nametbl”, tres también lo consideran el programa mejor entendido, de los ocho restantes: tres consideran que el programa mejor entendido es “Cmdline” y los otros cinco “Ntree”. Finalmente, de catorce personas que dicen que el programa con faltas más difíciles es “Ntree”, una también lo considera el programa mejor entendido, de los trece restantes: seis consideran que el programa mejor entendido es “Cmdline” y los otros siete “Nametbl”.

Para la iteración 3, de dieciséis personas que dicen que el programa con faltas más difíciles es “Cmdline”, ninguna lo consideran el programa mejor entendido, por lo tanto los dieciséis sujetos se distribuyen así: cinco consideran que el mejor entendido es “Nametbl” y los otros once “Ntree”. De quince personas que dicen que el programa con faltas más difíciles es “Nametbl”, una también lo considera el programa mejor entendido, de los catorce restantes: diez consideran que el programa mejor entendido es “Cmdline” y los otros cuatro “Ntree”. Finalmente, de catorce personas que dicen que el programa con faltas más difíciles es “Ntree”, dos también lo consideran el programa mejor entendido, de los doce restantes: siete consideran que el programa mejor entendido es “Cmdline” y los otros cinco “Nametbl”.

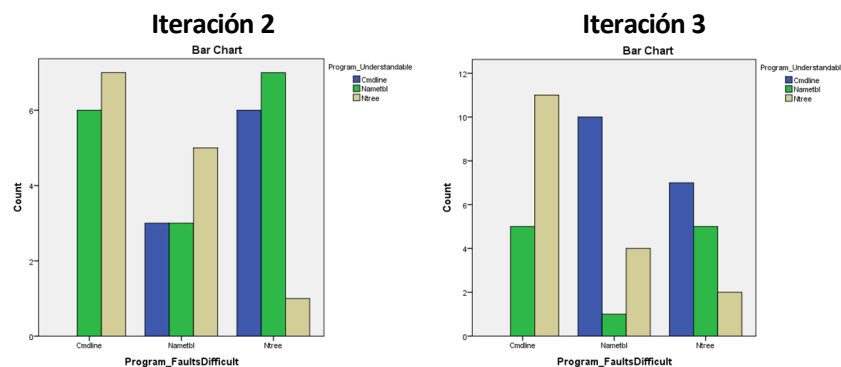


Figura 5.35. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.73y en la Tabla 5.74 para las iteraciones 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP3			Total
			Cmdblne	Nametbl	Ntree	
PP2	Cmdblne	Count	0	6	7	13
		% of Total	0,0%	15,8%	18,4%	34,2%
	Nametbl	Count	3	3	5	11
		% of Total	7,9%	7,9%	13,2%	28,9%
	Ntree	Count	6	7	1	14
		% of Total	15,8%	18,4%	2,6%	36,8%
Total		Count	9	16	13	38
		% of Total	23,7%	42,1%	34,2%	100,0%

Tabla 5.73. Tabla de Contingencia de la iteración 2

			PP3			Total
			Cmdline	Nametbl	Ntree	
PP2	Cmdline	Count	0	5	11	16
		% of Total	0,0%	11,1%	24,4%	35,6%
	Nametbl	Count	10	1	4	15
		% of Total	22,2%	2,2%	8,9%	33,3%
	Ntree	Count	7	5	2	14
		% of Total	15,6%	11,1%	4,4%	31,1%
Total		Count	17	11	17	45
		% of Total	37,8%	24,4%	37,8%	100,0%

Tabla 5.74. Tabla de Contingencia de la iteración 3

5.4.3.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 2 es del 10,5%, es decir muy pocos sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 6,6%, es decir muy pocos sujetos están de acuerdo. Se esperaban estos porcentajes tan bajos, puesto que las preguntas pueden considerarse contradictorias.

En la Tabla 6.75, se puede observar que, tanto para la iteración 2 como para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa es negativo se interpretará el valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 2	-,333	,073	-2,980	,003
Iteración 3	-,400	,060	-3,832	,000

Tabla 5.75. Índice de Concordancia Kappa

Debido a que en las dos iteraciones hay un desacuerdo discreto, se puede concluir que, como era de esperar, los sujetos no identifican el programa que tiene las faltas más difíciles con el programa que han entendido mejor.

5.4.4. ¿Qué programa es más fácil? / ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre el programa que los sujetos consideran que es el más fácil y el programa que creen que tiene las faltas más fáciles. La segunda pregunta aparece únicamente en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.4.1. Estudio del Acuerdo Después-Después

En la Figura 5.36, para la iteración 3, de veinte personas que dicen que el programa más fácil es "Cmdline", quince también lo consideran el programa con faltas más fáciles, de los cinco restantes:

cuatro consideran que el programa con las faltas más fáciles es “Nametbl” y otro “Ntree”. De once personas que dicen que el programa más fácil es “Nametbl”, ocho también lo consideran el programa con faltas más fáciles, de los tres restantes: uno considera que el programa con las faltas más fáciles es “Cmdline” y los otros dos “Ntree”. Finalmente, de catorce personas que dicen que el programa más fácil es “Ntree”, seis también lo considera el programa con faltas más fáciles, de los ocho restantes: cuatro consideran que el programa con las faltas más fáciles es “Cmdline” y los otros cuatro “Nametbl”.

Estos resultados se pueden contrastar en la Tabla 5.76 para la iteración 3. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

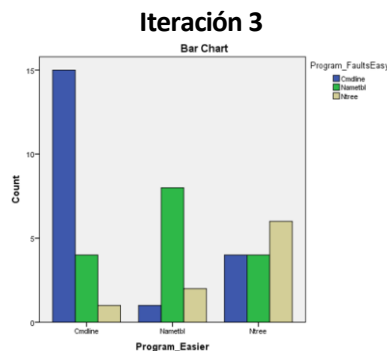


Figura 5.36. Acuerdo entre las respuestas

			PP1			Total
			Cmdline	Nametbl	Ntree	
PP1	Cmdline	Count	15	4	1	20
		% of Total	33,3%	8,9%	2,2%	44,4%
	Nametbl	Count	1	8	2	11
		% of Total	2,2%	17,8%	4,4%	24,4%
	Ntree	Count	4	4	6	14
		% of Total	8,9%	8,9%	13,3%	31,1%
Total		Count	20	16	9	45
		% of Total	44,4%	35,6%	20,0%	100,0%

Tabla 5.76. Tabla de Contingencia de la iteración 3

5.4.4.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 64,4%, es decir la mayoría de los sujetos están de acuerdo.

En la Tabla 6.77, se puede observar que, para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el resultado de Kappa muestra que acuerdo es moderado ($>0,41 - 0,60$), se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	,456	,105	4,387	,000

Tabla 5.77. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un desacuerdo moderado, se puede concluir que, los sujetos identifican el programa más fácil con el programa que tiene las faltas más fáciles. La relación entre las preguntas es intermedia.

5.4.5. ¿Qué programa tiene las faltas más difíciles? / ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre el programa que los sujetos consideran que tiene las faltas más difíciles y el programa que creen que tiene las faltas más fáciles. La segunda pregunta aparece únicamente en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.5.1. Estudio del Acuerdo Después – Después

En la Figura 5.37, para la iteración 3, de diecisiete personas que dicen que el programa con las faltas más difíciles es “Cmdline”, una también lo considera el programa con faltas más fáciles, de los dieciséis restantes: doce consideran que el programa con las faltas más fáciles es “Nametbl” y los otros cuatro “Ntree”. De quince personas que dicen que el programa con las faltas más difíciles es “Nametbl”, ninguna lo considera el programa con faltas más fáciles, por lo tanto los quince sujetos se distribuyen así: diez consideran que el programa con las faltas más fáciles es “Cmdline” y los otros cinco “Ntree”. Finalmente, de catorce personas que dicen que el programa con las faltas más difíciles es “Ntree”, ninguna lo considera el programa con faltas más fáciles, por lo tanto los catorce sujetos se distribuyen así: nueve consideran que el programa con las faltas más fáciles es “Cmdline” y los otros cinco “Nametbl”.

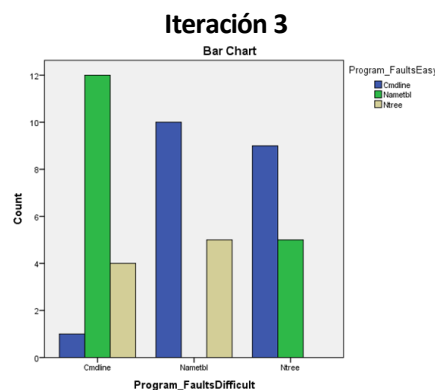


Figura 5.37. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.78 para la iteración 3. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP1			Total
			Cmdline	Nametbl	Ntree	
PP2	Cmdline	Count	1	12	4	17
		% of Total	2,2%	26,1%	8,7%	37,0%
	Nametbl	Count	10	0	5	15
		% of Total	21,7%	0,0%	10,9%	32,6%
	Ntree	Count	9	5	0	14
		% of Total	19,6%	10,9%	0,0%	30,4%
Total		Count	20	17	9	46
		% of Total	43,5%	37,0%	19,6%	100,0%

Tabla 5.78. Tabla de Contingencia de la iteración 3

5.4.5.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 2,2%, es decir muy pocos sujetos están de acuerdo. Se esperaba este resultado puesto que las preguntas son contradictorias.

En la Tabla 6.79, se puede observar que, para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa es negativo se interpretará el valor como desacuerdo, es decir, existe un desacuerdo moderado ($>0,41 - 0,60$), se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	-,484	,040	-4,677	,000

Tabla 5.79. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un desacuerdo moderado, se puede concluir que, como era de esperar, los sujetos no identifican el programa que tiene las faltas más difíciles con el programa que tiene las faltas más fáciles.

5.4.6. ¿Qué programa has entendido mejor? / ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre el programa que los sujetos consideran que es el más entendible y el programa que creen que tiene las faltas más fáciles. La segunda pregunta aparece únicamente en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la única combinación posible: Después-Después.

5.4.6.1. Estudio del Acuerdo Después – Después

En la Figura 5.38, para la iteración 3, de diecisiete personas que dicen que el programa más entendible es “Cmdline”, catorce también lo considera el programa con faltas más fáciles, de los tres restantes: dos consideran que el programa con las faltas más fáciles es “Nametbl” y otro “Ntree”. De once personas que dicen que el programa más entendible es “Nametbl”, ocho también lo

consideran el programa con faltas más fáciles y los tres restantes consideran que el programa con las faltas más fáciles es “Cmdline”. Finalmente, de diecisiete personas que dicen que el programa más entendible es “Ntree”, ocho también lo considera el programa con faltas más fáciles, de los nueve restantes: tres consideran que el programa con las faltas más fáciles es “Cmdline” y los otros seis “Nametbl”.

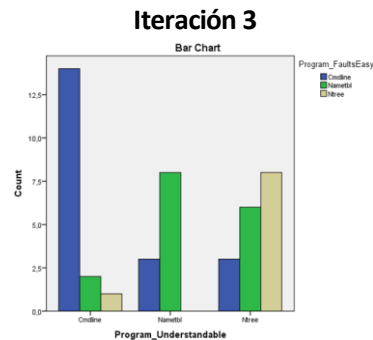


Figura 5.38. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 5.80 para la iteración 3. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP1			Total
			Cmdline	Nametbl	Ntree	
PP3	Cmdline	Count	14	2	1	17
		% of Total	31,1%	4,4%	2,2%	37,8%
	Nametbl	Count	3	8	0	11
		% of Total	6,7%	17,8%	0,0%	24,4%
	Ntree	Count	3	6	8	17
		% of Total	6,7%	13,3%	17,8%	37,8%
Total		Count	20	16	9	45
		% of Total	44,4%	35,6%	20,0%	100,0%

Tabla 5.80. Tabla de Contingencia de la iteración 3

5.4.6.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 66,7%, es decir la mayoría de los sujetos están de acuerdo.

En la Tabla 6.81, se puede observar que, para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el resultado de Kappa muestra que existe un acuerdo moderado ($> 0,41 - 0,60$), se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	,502	,100	4,975	,000

Tabla 5.81. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un acuerdo moderado, se puede concluir que los sujetos identifican el programa que han entendido mejor con el programa que tiene las faltas más fáciles. La relación entre las preguntas es intermedia.

CAPÍTULO 6. COMPARANDO LAS PREFERENCIAS CON LA REALIDAD

El objetivo de esta sección es responder a la pregunta: ¿sirven las preferencias para tomar decisiones sobre testing? Concretamente sobre la efectividad de las técnicas de testing.

Para ello, se estudia si ha habido concordancias para cada sujeto entre lo que prefieren y los resultados cuantitativos del estudio empírico. Esto se hace comparando las siguientes cosas.

Resultados Sobre Preferencias	Resultados Cuantitativos
Técnica Preferida	% defectos encontrados
Técnica más fácil de entender	
Técnica más fácil de aplicar	
Programa más fácil	
Programa mejor entendido	
Programa con faltas más fáciles	
Programa con faltas más difíciles	

Tabla 6.1 Comparación Preferencias con Realidad

A partir de los resultados se puede concluir si las preferencias de los sujetos funcionan o no funcionan en cada caso.

6.1. Preferencias sobre las Técnicas

6.1.1. ¿Qué técnica te gusta más?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y la técnica que les gusta más a los sujetos. Esta pregunta aparece en las tres iteraciones y se van a evaluar las respuestas que dieron los sujetos tanto antes como después de realizar el estudio empírico.

6.1.1.1. Respuestas Antes

6.1.1.1.1. Iteración 1

En la Tabla 6.2, se puede observar que para el caso de la técnica “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que la que mejor aplican es “Estructural”, finalmente en el caso de “Estructural” la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	38,0967	21,81852	3
	Structural	85,7100	,00000	3
	Functional	74,6033	22,50440	3
Structural	Review	53,0621	31,90691	14
	Structural	77,2086	24,77310	14
	Functional	81,6307	17,30644	14
Functional	Review	55,7130	8,10814	10
	Structural	52,8570	39,29588	10
	Functional	87,8550	10,19143	10

Tabla 6.2 Estadísticos Descriptivos

En la Tabla 6.3, se observa que se no puede asumir que las matrices de covarianzas sean iguales ($\text{sig} \leq 0.05$).

Box's M ^a	21,260
F	2,988
df1	6
df2	2550,816
Sig.	,007

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

Tabla 6.3. Test de igualdad de matrices de covarianzas

Aunque no se puede evaluar el test de Mauchly, en la Tabla 6.4, se puede observar que hubiese sido posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,817	4,652	2	,098	,845	,978	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.4. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.5

como en la Tabla 6.6, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,446	9,253b	2,000	23,000	,001
	Wilks' Lambda	,554	9,253b	2,000	23,000	,001
	Hotelling's Trace	,805	9,253b	2,000	23,000	,001
	Roy's Largest Root	,805	9,253b	2,000	23,000	,001
AppliedTec * PT1(Before)	Pillai's Trace	,232	1,573	4,000	48,000	,197
	Wilks' Lambda	,770	1,606b	4,000	46,000	,189
	Hotelling's Trace	,297	1,631	4,000	44,000	,183
	Roy's Largest Root	,289	3,467c	2,000	24,000	,048

a. Design: Intercept + Before_PreferredTec

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.5. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	9906,039	1,690	5859,987	7,417	,003
	Huynh-Feldt	9906,039	1,956	5063,881	7,417	,002
	Lower-bound	9906,039	1,000	9906,039	7,417	,012
AppliedTec * PT1(Before)	Greenhouse-Geisser	5079,122	3,381	1502,295	1,901	,138
	Huynh-Feldt	5079,122	3,912	1298,201	1,901	,127
	Lower-bound	5079,122	2,000	2539,561	1,901	,171
Error(AppliedTec)	Greenhouse-Geisser	32055,239	40,571	790,104		
	Huynh-Feldt	32055,239	46,949	682,765		
	Lower-bound	32055,239	24,000	1335,635		

Tabla 6.6. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$), sin embargo, Roy's Largest Root es significativo en la interacción ($\text{sig} \leq 0.05$).

En la Tabla 6.7, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	243104,119	1	243104,119	558,848	,000
PT1(Before)	509,824	2	254,912	,586	,564
Error	10440,231	24	435,010		

Tabla 6.7. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras

- Las preferencias de los sujetos sobre una técnica no hacen que se detecten más faltas
- Las preferencias sirven para predecir. Por lo tanto, pueden garantizar que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.8, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y la más baja los que aplicaron “Revisión”. Esto se puede observar gráficamente en la Figura 6.1.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	48,957	5,877	36,828	61,086
Structural	71,925	7,150	57,169	86,682
Functional	81,363	3,695	73,738	88,988

Tabla 6.8. Medias estimadas

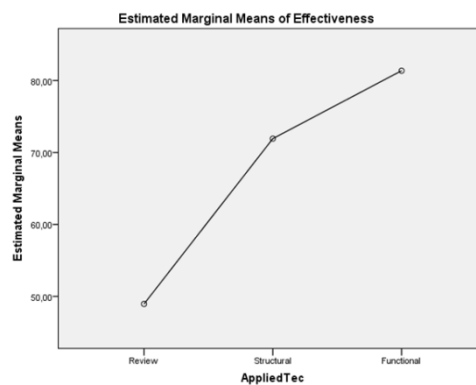


Figura 6.1 Técnica aplicada

Finalmente, en la Tabla 6.9, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Además, no existen diferencias significativas entre “Revisión” y “Estructural” y por último, tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
Review	Structural	-22,968	10,335	,108	-49,566	3,630
	Functional	-32,406*	7,499	,001	-51,706	-13,106
Structural	Review	22,968	10,335	,108	-3,630	49,566
	Functional	-9,438	7,854	,724	-29,651	10,775
Functional	Review	32,406*	7,499	,001	13,106	51,706
	Structural	9,438	7,854	,724	-10,775	29,651

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.9. Comparación por pares

Con respecto a la interacción de la técnica aplicada y la técnica que más les gusta, en la Tabla 6.10 se puede observar que de los que responden “Funcional” coincide con la técnica con la que han

encontrado más faltas, cuando responden “Revisión” encuentran más faltas en “Estructural”, finalmente, cuando responden “Estructural” encuentran más faltas en “Funcional”.

PT1(Before)	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	38,097	14,327	8,528	67,666
	Structural	85,710	17,431	49,735	121,685
	Functional	74,603	9,007	56,013	93,193
Structural	Review	53,062	6,632	39,374	66,750
	Structural	77,209	8,069	60,555	93,862
	Functional	81,631	4,170	73,025	90,236
Functional	Review	55,713	7,847	39,517	71,909
	Structural	52,857	9,547	33,153	72,561
	Functional	87,855	4,933	77,673	98,037

Tabla 6.10. Medias estimadas

En la Figura 6.2 se puede observar, que cuando los sujetos responden “Revisión” encuentran más faltas cuando aplican “Estructural”, aunque no hay mucha diferencia con “Funcional” y cuando aplican “Revisión” encuentran menos faltas. Cuando responden “Estructural” encuentran más faltas cuando aplican la misma, aunque no hay mucha diferencia con “Funcional”, y cuando aplican “Revisión” encuentran menos faltas. Finalmente, cuando responden “Funcional” encuentran más faltas cuando aplican la misma y encuentran menos faltas cuando aplican “Estructural”, aunque no hay mucha diferencia con “Revisión”.

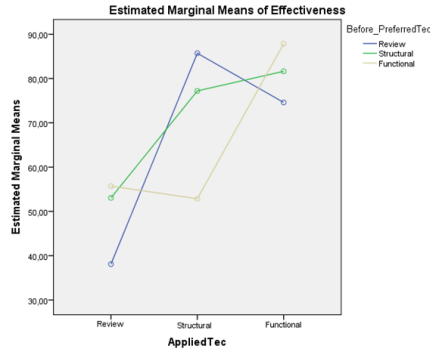


Figura 6.2 Técnica aplicada * Técnica Preferida (Antes)

Finalmente, en la Tabla 6.11, se puede comprobar que de los que dicen que la técnica que les gusta más es “Revisión”, no hay diferencias significativas entre “Revisión” y “Estructural”. Además entre “Revisión” y “Funcional” tampoco hay diferencias significativas. Finalmente tampoco hay diferencias significativas entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

Por otro lado de los que dicen que la técnica más efectiva es “Estructural”, aplican significativamente mejor “Funcional” que “Revisión”. Además no hay diferencias significativas cuando aplican “Revisión” y “Estructural”. Finalmente, entre “Estructural” y “Funcional” tampoco hay diferencias. Por lo tanto parece que no coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que la técnica más efectiva es “Funcional”, aplican significativamente mejor la misma que “Revisión” y “Estructural”. Además no hay diferencias significativas cuando aplican “Revisión” y “Estructural”. Por lo tanto coincide lo que responden los sujetos con lo que aplican.

PT1(Before)	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
Review	Review	Structural	-47,613	25,195	,213	-112,457	17,231
		Functional	-36,507	18,282	,172	-83,558	10,544
	Structural	Review	47,613	25,195	,213	-17,231	112,457
		Functional	11,107	19,147	1,000	-38,170	60,384
	Functional	Review	36,507	18,282	,172	-10,544	83,558
		Structural	-11,107	19,147	1,000	-60,384	38,170
Structural	Review	Structural	-24,146	11,663	,148	-54,163	5,870
		Functional	-28,569*	8,463	,008	-50,349	-6,788
	Structural	Review	24,146	11,663	,148	-5,870	54,163
		Functional	-4,422	8,863	1,000	-27,233	18,389
	Functional	Review	28,569*	8,463	,008	6,788	50,349
		Structural	4,422	8,863	1,000	-18,389	27,233
Functional	Review	Structural	2,856	13,800	1,000	-32,660	38,372
		Functional	-32,142*	10,013	,011	-57,913	-6,371
	Structural	Review	-2,856	13,800	1,000	-38,372	32,660
		Functional	-34,998*	10,487	,008	-61,988	-8,008
	Functional	Review	32,142*	10,013	,011	6,371	57,913
		Structural	34,998*	10,487	,008	8,008	61,988

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.11. Comparación por pares

6.1.1.1.2. Iteración 2

En la Tabla 6.12, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	54,1667	17,25164	8
	Structural	35,4167	20,77372	8
	Functional	50,0000	15,43033	8
Structural	Review	60,0000	19,00292	5
	Structural	50,0000	16,66667	5
	Functional	33,3333	11,78511	5
Functional	Review	55,3763	22,92786	31
	Structural	50,0000	14,90712	31
	Functional	48,9246	18,72593	31

Tabla 6.12. Estadísticos Descriptivos

En la Tabla 6.13, se puede observar que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	24,003
F	1,527
df1	12
df2	631,986
Sig.	,110

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

Tabla 6.13. Test de igualdad de matrices de covarianzas

En la Tabla 6.14, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,959	1,682	2	,431	,960	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.14. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.15, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que consideran la más efectiva, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	2393,004	2	1196,502	3,608	,031
AppliedTec * PT1(Before)	Sphericity Assumed	2069,443	4	517,361	1,560	,193
Error(AppliedTec)	Sphericity Assumed	27189,946	82	331,585		

Tabla 6.15. Test univariado de efectos intra-sujetos

En la Tabla 6.16, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	178357,774	1	178357,774	467,290	,000
PT1(Before)	552,472	2	276,236	,724	,491
Error	15649,091	41	381,685		

Tabla 6.16. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hacen que se detecten más faltas

- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.17, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Funcional” y “Estructural”, aunque no hay mucha diferencia, la técnica con menor media es “Funcional”. Esto se puede observar gráficamente en la Figura 6.3.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	56,514	4,323	47,785	65,244
Structural	45,139	3,233	38,609	51,669
Functional	44,086	3,512	36,993	51,179

Tabla 6.17. Medias estimadas

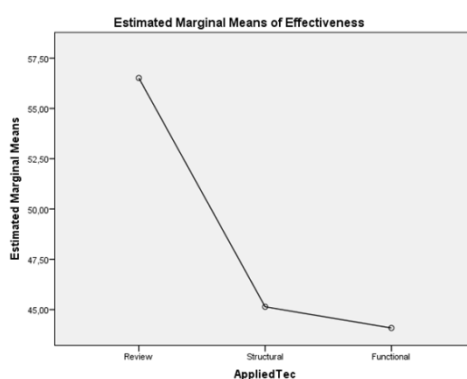


Figura 6.3. Técnica aplicada

Finalmente, en la Tabla 6.18, se puede comprobar que no hay diferencias significativas entre “Revisión” y “Estructural”, ni tampoco entre “Revisión” y “Funcional”, finalmente entre “Estructural” y “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	11,375	5,259	,109	-1,753	24,504
	Functional	12,428	5,490	,087	-1,275	26,132
Structural	Review	-11,375	5,259	,109	-24,504	1,753
	Functional	1,053	4,602	1,000	-10,434	12,540
Functional	Review	-12,428	5,490	,087	-26,132	1,275
	Structural	-1,053	4,602	1,000	-12,540	10,434

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.18. Comparación por pares

6.1.1.1.3. Iteración 3

En la Tabla 6.19, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide

ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	51,6680	26,58590	10
	Structural	41,6660	22,56595	10
	Functional	41,6660	11,78511	10
Structural	Review	50,0000	13,61100	13
	Structural	42,3069	23,19014	13
	Functional	38,4600	10,50746	13
Functional	Review	57,0174	23,11824	19
	Structural	41,2284	21,77766	19
	Functional	42,1053	15,08067	19

Tabla 6.19. Estadísticos Descriptivos

En la Tabla 6.20, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	15,137
F	1,105
df1	12
df2	4203,910
Sig.	,351

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

Tabla 6.20. Test de igualdad de matrices de covarianzas

En la Tabla 6.21, se puede observar que no es posible asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados el valor de Huyn-Feldt.

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,851	6,129	2	,047	,870	,954	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.21. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.22 como en la Tabla 6.23, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,205	4,910b	2,000	38,000	,013
	Wilks' Lambda	,795	4,910b	2,000	38,000	,013
	Hotelling's Trace	,258	4,910b	2,000	38,000	,013
	Roy's Largest Root	,258	4,910b	2,000	38,000	,013
AppliedTec * PT1(Before)	Pillai's Trace	,017	,168	4,000	78,000	,954
	Wilks' Lambda	,983	,164b	4,000	76,000	,956
	Hotelling's Trace	,017	,160	4,000	74,000	,958
	Roy's Largest Root	,014	,273c	2,000	39,000	,763

a. Design: Intercept + Before_PreferredTec

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.22. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3570,541	1,741	2051,193	4,870	,014
	Huynh-Feldt	3570,541	1,909	1870,845	4,870	,011
	Lower-bound	3570,541	1,000	3570,541	4,870	,033
AppliedTec * PT1(Before)	Greenhouse-Geisser	305,742	3,481	87,821	,209	,914
	Huynh-Feldt	305,742	3,817	80,099	,209	,927
	Lower-bound	305,742	2,000	152,871	,209	,813
Error(AppliedTec)	Greenhouse-Geisser	28592,525	67,888	421,173		
	Huynh-Feldt	28592,525	74,432	384,142		
	Lower-bound	28592,525	39,000	733,142		

Tabla 6.23. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$).

En la Tabla 6.24, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$), es decir, la opinión de los sujetos es la misma para las tres técnicas.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	239495,339	1	239495,339	589,154	,000
PT1(Before)	241,732	2	120,866	,297	,744
Error	15853,775	39	406,507		

Tabla 6.24. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.25, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Funcional” y “Estructural”, aunque no hay mucha diferencia, la técnica con menor media es “Funcional”. Esto se puede observar gráficamente en la Figura 6.4.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	52,895	3,450	45,916	59,874
Structural	41,734	3,578	34,497	48,971
Functional	40,744	2,088	36,520	44,968

Tabla 6.25. Medias estimadas

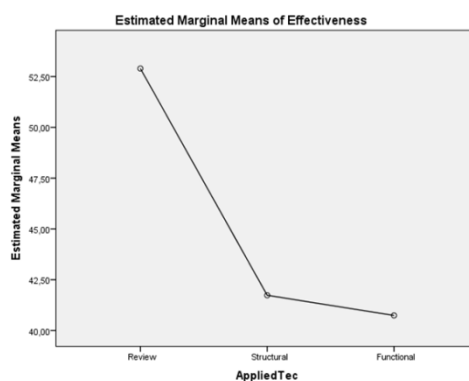


Figura 6.4. Técnica aplicada

Finalmente, en la Tabla 6.26, se puede comprobar que “Revisión” es significativamente mejor que “Funcional”. Por otro lado, no hay significativas entre “Revisión” y “Estructural”, finalmente tampoco hay diferencias entre “Estructural” y “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	11,161	5,089	,103	-1,571	23,893
	Functional	12,151*	3,835	,009	2,557	21,746
Structural	Review	-11,161	5,089	,103	-23,893	1,571
	Functional	,990	3,935	1,000	-8,855	10,835
Functional	Review	-12,151*	3,835	,009	-21,746	-2,557
	Structural	-,990	3,935	1,000	-10,835	8,855

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.26. Comparación por pares

6.1.1.1.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.27, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas en la misma que en “Revisión”. Con respecto a la interacción de la técnica aplicada y la técnica preferida, cuando los sujetos responden “Revisión” encuentran el

mismo número de faltas al aplicar las tres técnicas, cuando responden “Estructural” aplican mejor “Funcional” que “Revisión” pero no que “Estructural”; finalmente cuando responden “Funcional” aplican mejor “Funcional” que “Revisión” y “Estructural”. Además no influyen las preferencias.

En el caso de la iteración 2, la técnica aplicada sale significativa, sin embargo, no hay ninguna técnica que se haya aplicado mejor que las otras. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Finalmente, en la iteración 3 cuando los sujetos aplican la técnica “Revisión” encuentran más faltas que en “Funcional”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Debido a que en dos iteraciones hay técnicas que detectan más faltas y estas no son la misma, se podría decir, que realmente no es la técnica la que determina este comportamiento. Es decir, al parecer existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que influyen al aplicar las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,001	F>R	,564	N/A	,197	R R=E=F
		R=E				F>R
		E=F				R=E
						E=F
Iteración 2	,031	R=E=F	,491	N/A	,193	N/A
Iteración 3	,013	R>F	,744	N/A	,954	N/A
		R=E				
		E=F				

Tabla 6.27. Resumen

6.1.1.2. Respuestas Después

6.1.1.2.1. Iteración 1

En la Tabla 6.28, se puede observar que para el caso de las técnicas “Estructural” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que la que mejor aplican es “Estructural”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	14,2900	.	1
	Structural	85,7100	.	1
	Functional	57,1400	.	1
Structural	Review	57,1443	34,00440	7
	Structural	81,6300	13,58825	7
	Functional	71,4271	18,64592	7
Functional	Review	52,3806	20,20176	18
	Structural	62,4328	36,54443	18
	Functional	88,2256	10,40935	18

Tabla 6.28. Estadísticos Descriptivos

En la Tabla 6.29, se observa que se no puede asumir que las matrices de covarianzas sean iguales ($\text{sig} \leq 0.05$).

Box's M ^a	21,260
F	2,988
df1	6
df2	2550,816
Sig.	,005

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

Tabla 6.29. Test de igualdad de matrices de covarianzas

Aunque no se puede evaluar el test de Mauchly, en la Tabla 6.30, se observa que tampoco es posible asumir esfericidad ($\text{sig} \leq 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,718	7,283	2	,026	,780	,899	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.30. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.31 como en la Tabla 6.32, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,252	3,707b	2,000	22,000	,041
	Wilks' Lambda	,748	3,707b	2,000	22,000	,041
	Hotelling's Trace	,337	3,707b	2,000	22,000	,041
	Roy's Largest Root	,337	3,707b	2,000	22,000	,041
AppliedTec * PT1(After)	Pillai's Trace	,334	2,302	4,000	46,000	,073
	Wilks' Lambda	,679	2,354b	4,000	44,000	,068
	Hotelling's Trace	,456	2,394	4,000	42,000	,066
	Roy's Largest Root	,413	4,747c	2,000	23,000	,019

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.31. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	5574,165	1,560	3572,529	4,008	,036
	Huynh-Feldt	5574,165	1,799	3098,691	4,008	,029
	Lower-bound	5574,165	1,000	5574,165	4,008	,057
AppliedTec * PT1(After)	Greenhouse-Geisser	5044,191	3,121	1616,432	1,813	,160
	Huynh-Feldt	5044,191	3,598	1402,039	1,813	,150
	Lower-bound	5044,191	2,000	2522,095	1,813	,186
Error(AppliedTec)	Greenhouse-Geisser	31991,244	35,887	891,455		
	Huynh-Feldt	31991,244	41,374	773,218		
	Lower-bound	31991,244	23,000	1390,924		

Tabla 6.32. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$), sin embargo, Lower-bound no es significativo para la técnica aplicada ($\text{sig} > 0.05$), y Roy's Largest Root es significativo en la interacción ($\text{sig} \leq 0.05$).

En la Tabla 6.33, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$), es decir, la opinión de los sujetos es la misma para las tres técnicas.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	90490,170	1	90490,170	216,264	,000
PT1(After)	821,710	2	410,855	,982	,390
Error	9623,762	23	418,424		

Tabla 6.33. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias sirven para predecir. Por lo tanto, pueden garantizar que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.34, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y la más baja los que aplicaron “Revisión”. Esto se puede observar gráficamente en la Figura 6.5.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	41,272	8,963	22,731	59,813
Structural	76,591	11,741	52,303	100,879
Functional	72,264	4,769	62,399	82,129

Tabla 6.34. Medias estimadas

Finalmente, en la Tabla 6.35, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Además, no existen diferencias significativas entre “Revisión” y “Estructural” y por último, tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

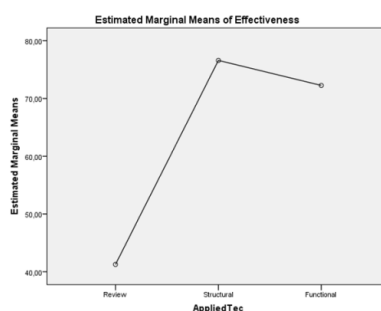


Figura 6.5. Técnica aplicada

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	-35,319	16,812	,140	-78,727	8,089
	Functional	-30,993*	11,180	,033	-59,861	-2,125
Structural	Review	35,319	16,812	,140	-8,089	78,727
	Functional	4,327	12,165	1,000	-27,085	35,738
Functional	Review	30,993*	11,180	,033	2,125	59,861
	Structural	-4,327	12,165	1,000	-35,738	27,085

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.35. Comparación por pares

Con respecto a la interacción de la técnica aplicada y la técnica que más les gusta, en la Tabla 6.36 se puede observar que de los que responden “Funcional” y “Estructural” coincide con la técnica con la que han encontrado más faltas, cuando responden “Revisión” encuentran más faltas en “Estructural”.

PT1(After)	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	14,290	24,562	-36,520	65,100
	Structural	85,710	32,176	19,150	152,270
	Functional	57,140	13,068	30,106	84,174
Structural	Review	57,144	9,284	37,940	76,349
	Structural	81,630	12,161	56,473	106,787
	Functional	71,427	4,939	61,209	81,645
Functional	Review	52,381	5,789	40,404	64,357
	Structural	62,433	7,584	46,744	78,121
	Functional	88,226	3,080	81,854	94,598

Tabla 6.36. Medias estimadas

En la Figura 6.6 se puede observar, que cuando los sujetos responden “Revisión” encuentran más faltas cuando aplican “Estructural”, aunque hay poca diferencia con “Funcional” y cuando aplican “Revisión” encuentran menos faltas. Cuando responden “Estructural” encuentran más faltas cuando aplican la misma, aunque no hay mucha diferencia entre las tres técnicas. Finalmente, cuando responden “Funcional” encuentran más faltas cuando aplican la misma y encuentran menos faltas cuando aplican “Estructural”, aunque no hay mucha diferencia con “Revisión”.

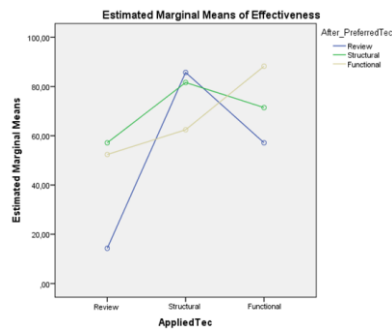


Figura 6.6 Técnica aplicada * Técnica Preferida (Después)

Finalmente, en la Tabla 6.37, se puede comprobar que de los que dicen que la técnica que les gusta más es “Revisión”, no hay diferencias significativas entre “Revisión” y “Estructural”. Además entre “Revisión” y “Funcional” tampoco hay diferencias significativas. Finalmente tampoco hay diferencias significativas entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

Por otro lado de los que dicen que la técnica más efectiva es “Estructural”, no hay diferencias significativas entre “Revisión” y “Estructural”. Además entre “Revisión” y “Funcional” tampoco hay diferencias significativas. Finalmente tampoco hay diferencias significativas entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que la técnica más efectiva es “Funcional”, aplican significativamente mejor la misma que “Revisión” y “Estructural”. Y no hay diferencias significativas entre “Revisión” y “Estructural”. Por lo tanto coincide lo que responden con lo que aplican.

PT1(After)	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Review	Review	Structural	-71,420	46,071	,404	-190,377	47,537
		Functional	-42,850	30,639	,526	-121,961	36,261
	Structural	Review	71,420	46,071	,404	-47,537	190,377
		Functional	28,570	33,339	1,000	-57,511	114,651
	Functional	Review	42,850	30,639	,526	-36,261	121,961
		Structural	-28,570	33,339	1,000	-114,651	57,511
Structural	Review	Structural	-24,486	17,413	,519	-69,447	20,476
		Functional	-14,283	11,581	,690	-44,184	15,618
	Structural	Review	24,486	17,413	,519	-20,476	69,447
		Functional	10,203	12,601	1,000	-22,333	42,738
	Functional	Review	14,283	11,581	,690	-15,618	44,184
		Structural	-10,203	12,601	1,000	-42,738	22,333
Functional	Review	Structural	-10,052	10,859	1,000	-38,091	17,986
		Functional	-35,845*	7,222	,000	-54,492	-17,198
	Structural	Review	10,052	10,859	1,000	-17,986	38,091
		Functional	-25,793*	7,858	,010	-46,082	-5,503
	Functional	Review	35,845*	7,222	,000	17,198	54,492
		Structural	25,793*	7,858	,010	5,503	46,082

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.37. Comparación por pares

6.1.1.2.2. Iteración 2

En la Tabla 6.38, se puede observar que para el caso de la técnica “Revisión” y “Estructural” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Funcional” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	59,2593	20,60055	9
	Structural	50,0000	14,43376	9
	Functional	49,9996	14,43424	9
Structural	Review	50,0000	28,86751	3
	Structural	50,0000	16,66667	3
	Functional	22,2222	25,45875	3
Functional	Review	50,6410	23,79578	26
	Structural	44,2308	18,82215	26
	Functional	48,7179	16,27751	26

Tabla 6.38. Estadísticos Descriptivos

En la Tabla 6.39, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	1,931
F	,273
df1	6
df2	1359,451
Sig.	,950

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

Tabla 6.39. Test de igualdad de matrices de covarianzas

En la Tabla 6.40, se observa que se asume esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,949	1,784	2	,410	,951	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.40. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.41, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que creen que aplicaron mejor, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	1591,699	2	795,849	2,106	,129
AppliedTec * PT1(After)	Sphericity Assumed	1650,516	4	412,629	1,092	,367
Error(AppliedTec)	Sphericity Assumed	26453,910	70	377,913		

Tabla 6.41. Test univariado de efectos intra-sujetos

En la Tabla 6.42, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	124720,779	1	124720,779	325,039	,000
PT1(After)	1141,237	2	570,618	1,487	,240
Error	13429,866	35	383,710		

Tabla 6.42. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

6.1.1.2.3. Iteración 3

En la Tabla 6.43, se puede observar que para el caso de la técnica “Revisión” y “Estructural” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Funcional” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT1(After)	AppliedTec	Mean	Std. Desviation	N
Review	Review	63,0964	19,80631	14
	Structural	41,6657	20,41241	14
	Functional	41,6664	16,98453	14
Structural	Review	50,0000	13,61100	4
	Structural	54,1675	15,96031	4
	Functional	33,3325	13,60692	4
Functional	Review	49,3826	21,42482	27
	Structural	39,5063	26,61256	27
	Functional	40,7400	13,34460	27

Tabla 6.43. Estadísticos Descriptivos

En la Tabla 6.44, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$). En la Tabla 6.45, se observa que no se asume esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados el valor de Huyn-Feldt.

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.46 como en la Tabla 6.47, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con

respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Box's M ^a	13,148
F	,801
df1	12
df2	308,977
Sig.	,650

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

Tabla 6.44. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,821	8,095	2	,017	,848	,922	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT1(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.45. Test de esfericidad de Mauchly

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,226	5,994b	2,000	41,000	,005
	Wilks' Lambda	,774	5,994b	2,000	41,000	,005
	Hotelling's Trace	,292	5,994b	2,000	41,000	,005
	Roy's Largest Root	,292	5,994b	2,000	41,000	,005
AppliedTec * After_PreferredTec	Pillai's Trace	,120	1,335	4,000	84,000	,264
	Wilks' Lambda	,884	1,304b	4,000	82,000	,275
	Hotelling's Trace	,127	1,274	4,000	80,000	,287
	Roy's Largest Root	,075	1,568c	2,000	42,000	,221

a. Design: Intercept + After_PreferredTec

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.46. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3073,630	1,696	1812,158	4,026	,028
	Huynh-Feldt	3073,630	1,844	1666,719	4,026	,025
	Lower-bound	3073,630	1,000	3073,630	4,026	,051
AppliedTec * PT1(After)	Greenhouse-Geisser	1896,149	3,392	558,968	1,242	,301
	Huynh-Feldt	1896,149	3,688	514,107	1,242	,301
	Lower-bound	1896,149	2,000	948,075	1,242	,299
Error(AppliedTec)	Greenhouse-Geisser	32067,503	71,237	450,153		
	Huynh-Feldt	32067,503	77,453	414,025		
	Lower-bound	32067,503	42,000	763,512		

Tabla 6.47. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$), sin embargo, Lower-bound no es significativo para la técnica aplicada ($\text{sig} > 0.05$).

En la Tabla 6.48, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	159038,756	1	159038,756	350,580	,000
PT1(After)	872,874	2	436,437	,962	,390
Error	19053,081	42	453,645		

Tabla 6.48. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.49, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y la más baja los que aplicaron “Funcional”. Esto se puede observar gráficamente en la Figura 6.7.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	54,160	4,084	45,917	62,402
Structural	45,113	4,829	35,367	54,859
Functional	38,580	2,911	32,705	44,454

Tabla 6.49. Medias estimadas

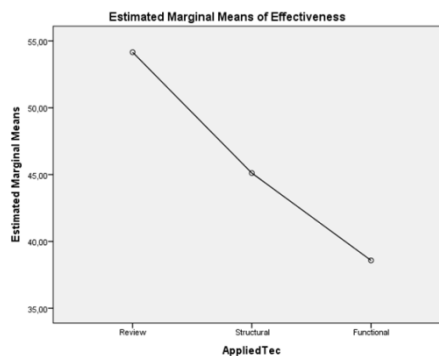


Figura 6.7. Técnica aplicada

Finalmente, en la Tabla 6.50, se puede comprobar que la técnica “Revisión” es significativamente mejor que la “Funcional”. Además, no existen diferencias significativas entre “Revisión” y “Estructural” y por último, tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	9,047	6,535	,521	-7,250	25,343
	Functional	15,580*	4,579	,004	4,161	26,999
Structural	Review	-9,047	6,535	,521	-25,343	7,250
	Functional	6,534	5,249	,660	-6,556	19,623
Functional	Review	-15,580*	4,579	,004	-26,999	-4,161
	Structural	-6,534	5,249	,660	-19,623	6,556

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.50. Comparación por pares

6.1.1.2.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.51, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas que en “Revisión”. Con respecto a la interacción de la técnica aplicada y la técnica preferida, cuando los sujetos responden “Revisión” y “Estructural” encuentran el mismo número de faltas con las tres técnicas, sin embargo, cuando responden “Funcional” encuentran más faltas con las misma que con “Revisión” y “Estructural”. Finalmente no influyen las preferencias que se tienen sobre las técnicas.

Finalmente, en la iteración 3 cuando los sujetos aplican la técnica “Revisión” encuentran más faltas que en “Funcional”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Debido a que en dos iteraciones hay técnicas que detectan más faltas y estas no son la misma, se podría decir, que realmente no es la técnica la que determina este comportamiento. Es decir, al parecer existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que influyen al aplicar las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,041	F>R	,390	N/A	,073	R
		R=E				E
		E=F				F
Iteración 2	,129	N/A	,240	N/A	,367	N/A
Iteración 3	,005	R>F	,390	N/A	,264	R=E=F
		R=E				R=E=F
		E=F				F>(R=E)

Tabla 6.51. Resumen

6.1.2. ¿Qué técnica te ha parecido más fácil de entender?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y la técnica que los sujetos consideran que es la más fácil de entender. Esta pregunta aparece en las tres iteraciones y únicamente se realizó después del estudio empírico.

6.1.2.1. Iteración 1

En la Tabla 6.52, se puede observar que para el caso de la técnica “Estructural” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

PT2	AppliedTec	Mean	Std. Deviation	N
Review	Review	63,4933	19,04688	9
	Structural	77,5111	17,53981	9
	Functional	78,8344	20,77154	9
Structural	Review	54,7600	29,15971	6
	Structural	61,5067	48,14909	6
	Functional	61,5067	48,14909	6
Functional	Review	47,6200	35,94715	3
	Structural	47,6200	41,24013	3
	Functional	90,4733	8,25034	3

Tabla 6.52. Estadísticos Descriptivos

En la Tabla 6.53, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	11,271
F	1,359
df1	6
df2	753,307
Sig.	,229

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

Tabla 6.53. Test de igualdad de matrices de covarianzas

En la Tabla 6.54, se observa que se puede asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,799	3,149	2	,207	,832	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.54. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.55, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la

interacción de la técnica que los sujetos aplicaron y la técnica que creen que es más fácil de entender, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	6917,542	2	3458,771	4,142	,026
AppliedTec * PT2	Sphericity Assumed	2273,920	4	568,480	,681	,611
Error(AppliedTec)	Sphericity Assumed	25053,150	30	835,105		

Tabla 6.55. Test univariado de efectos intra-sujetos

En la Tabla 6.56, se observa que las opiniones de la técnica más fácil de entender no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	200535,747	1	200535,747	478,150	,000
PT2	1020,952	2	510,476	1,217	,324
Error	6290,992	15	419,399		

Tabla 6.56. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica elegida, en la Tabla 6.57, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y entre los que aplicaron “Estructural” y “Revisión” la técnica con menor media es “Revisión”. Esto se puede observar gráficamente en la Figura 6.8.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	55,291	6,639	41,140	69,443
Structural	62,213	8,889	43,266	81,159
Functional	84,610	4,335	75,370	93,850

Tabla 6.57. Medias estimadas

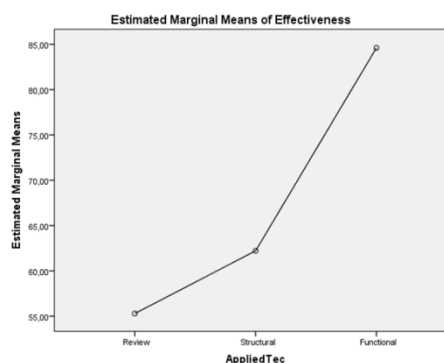


Figura 6.8. Técnica aplicada

Finalmente, en la Tabla 6.58, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	-6,921	12,813	1,000	-41,435	27,593
	Functional	-29,319*	9,205	,018	-54,114	-4,524
Structural	Review	6,921	12,813	1,000	-27,593	41,435
	Functional	-22,397	9,557	,100	-48,142	3,347
Functional	Review	29,319*	9,205	,018	4,524	54,114
	Structural	22,397	9,557	,100	-3,347	48,142

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.58. Comparación por pares

6.1.2.2. Iteración 2

En la Tabla 6.59, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT2	AppliedTec	Mean	Std. Deviation	N
Review	Review	58,3333	18,19881	14
	Structural	42,8571	22,37433	14
	Functional	46,4283	11,65517	14
Structural	Review	69,4444	12,54621	6
	Structural	44,4444	13,60828	6
	Functional	58,3333	13,94433	6
Functional	Review	50,6944	23,30182	24
	Structural	50,6944	13,44084	24
	Functional	45,1389	21,12716	24

Tabla 6.59. Estadísticos Descriptivos

En la Tabla 6.60, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	19,854
F	1,370
df1	12
df2	1026,757
Sig.	,174

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

Tabla 6.60. Test de igualdad de matrices de covarianzas

En la Tabla 6.61, se observa que se puede asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,954	1,869	2	,393	,956	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.61. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.62, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que creen que es la más fácil de entender, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	3093,559	2	1546,779	4,684	,012
AppliedTec * PT2	Sphericity Assumed	2178,131	4	544,533	1,649	,170
Error(AppliedTec)	Sphericity Assumed	27081,258	82	330,259		

Tabla 6.62. Test univariado de efectos intra-sujetos

En la Tabla 6.63, se observa que las opiniones de la técnica más fácil de entender no son significativas ($\text{sig} > 0.05$), es decir, la opinión de los sujetos es la misma para las tres técnicas.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	259148,777	1	259148,777	703,971	,000
PT2	1108,462	2	554,231	1,506	,234
Error	15093,101	41	368,124		

Tabla 6.63. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.64, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Estructural” y “Funcional” la técnica con menor media es “Estructural”. Esto se puede observar gráficamente en la Figura 6.9.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	59,491	3,651	52,118	66,864
Structural	45,999	2,964	40,012	51,985
Functional	49,967	3,140	43,626	56,308

Tabla 6.64. Medias estimadas

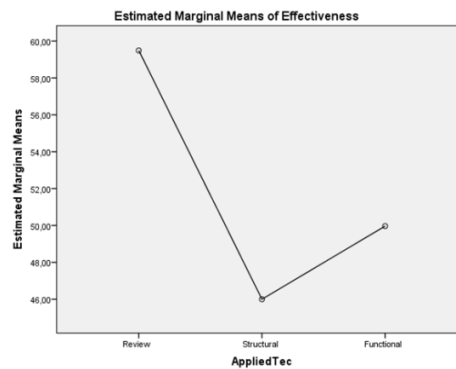


Figura 6.9. Técnica aplicada

Finalmente, en la Tabla 6.65, se puede comprobar que la técnica “Revisión” es significativamente mejor que la “Estructural”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Funcional”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	13,492*	4,412	,012	2,479	24,505
	Functional	9,524	4,975	,188	-2,895	21,943
Structural	Review	-13,492*	4,412	,012	-24,505	-2,479
	Functional	-3,968	4,169	1,000	-14,375	6,438
Functional	Review	-9,524	4,975	,188	-21,943	2,895
	Structural	3,968	4,169	1,000	-6,438	14,375

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.65. Comparación por pares

6.1.2.3. Iteración 3

En la Tabla 6.66, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT2	AppliedTec	Mean	Std. Desviation	N
Review	Review	54,3868	23,46737	19
	Structural	43,8595	26,18138	19
	Functional	46,4905	10,51216	19
Structural	Review	46,2967	21,69681	9
	Structural	42,5922	22,22146	9
	Functional	35,1844	13,02799	9
Functional	Review	56,4092	18,68175	13
	Structural	38,4615	15,78856	13
	Functional	37,1792	13,86658	13

Tabla 6.66. Estadísticos Descriptivos

En la Tabla 6.67, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	9,987
F	,722
df1	12
df2	3299,672
Sig.	,731

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

Tabla 6.67. Test de igualdad de matrices de covarianzas

En la Tabla 6.68, se observa que no se puede asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados el valor de Huyn-Feldt.

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,843	6,306	2	,043	,865	,950	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT2

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.68. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.69 como en la Tabla 6.70, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica más fácil de entender, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,223	5,296b	2,000	37,000	,009
	Wilks' Lambda	,777	5,296b	2,000	37,000	,009
	Hotelling's Trace	,286	5,296b	2,000	37,000	,009
	Roy's Largest Root	,286	5,296b	2,000	37,000	,009
AppliedTec * PT2	Pillai's Trace	,069	,683	4,000	76,000	,606
	Wilks' Lambda	,932	,666b	4,000	74,000	,618
	Hotelling's Trace	,072	,649	4,000	72,000	,629
	Roy's Largest Root	,049	,923c	2,000	38,000	,406

a. Design: Intercept + UnderstableTec

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.69. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3510,288	1,729	2030,193	4,818	,015
	Huynh-Feldt	3510,288	1,899	1848,167	4,818	,012
	Lower-bound	3510,288	1,000	3510,288	4,818	,034
AppliedTec * PT2	Greenhouse-Geisser	909,824	3,458	263,101	,624	,624
	Huynh-Feldt	909,824	3,799	239,511	,624	,639
	Lower-bound	909,824	2,000	454,912	,624	,541
Error(AppliedTec)	Greenhouse-Geisser	27685,865	65,704	421,375		
	Huynh-Feldt	27685,865	72,175	383,595		
	Lower-bound	27685,865	38,000	728,575		

Tabla 6.70. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$).

En la Tabla 6.71, se observa que las opiniones de la técnica que más les gusta a los sujetos no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	222561,608	1	222561,608	559,514	,000
PT2	977,351	2	488,676	1,229	,304
Error	15115,515	38	397,777		

Tabla 6.71. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, En la Tabla 6.72, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Estructural” y “Funcional” la técnica con menor media es “Funcional”. Esto se puede observar gráficamente en la Figura 6.10.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	52,364	3,546	45,186	59,542
Structural	41,638	3,683	34,181	49,094
Functional	39,618	1,995	35,580	43,656

Tabla 6.72. Medias estimadas

Finalmente, en la Tabla 6.73, se puede comprobar que la técnica “Revisión” es significativamente mejor que la “Funcional”. Por otro lado no existen diferencias significativas entre

la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

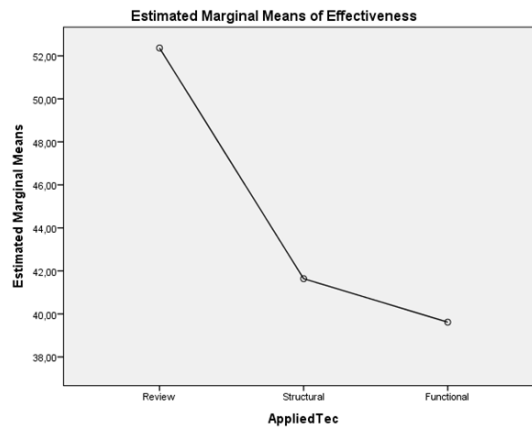


Figura 6.10. Técnica aplicada

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	10,727	5,211	,139	-2,324	23,777
	Funcional	12,746*	3,865	,006	3,065	22,427
Structural	Review	-10,727	5,211	,139	-23,777	2,324
	Funcional	2,020	4,044	1,000	-8,109	12,149
Funcional	Review	-12,746*	3,865	,006	-22,427	-3,065
	Structural	-2,020	4,044	1,000	-12,149	8,109

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.73. Comparación por pares

6.1.2.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.74, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas que en “Revisión”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Finalmente, en las iteraciones 2 y 3 cuando los sujetos aplican la técnica “Revisión” encuentran más faltas que en “Funcional”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Debido a que en dos de las tres iteraciones, una misma técnica es la que detecta más faltas, se puede concluir que, efectivamente, hay técnicas que detectan más faltas, sin embargo, no se puede descartar la influencia de otros factores, puesto que en la iteración 1 la técnica con la que se han encontrado más faltas es diferente.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,006	F>R	,324	N/A	,611	N/A
		R=E				
		E=F				
Iteración 2	,012	R>E	,234	N/A	,170	N/A
		R=F				
		E=F				
Iteración 3	,009	R>F	,304	N/A	,606	N/A
		R=E				
		E=F				

Tabla 6.74. Resumen

6.1.3. ¿Qué técnica te ha parecido más fácil de aplicar?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y la técnica que los sujetos consideran que es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones y se realizó tanto antes como después del estudio empírico.

6.1.3.1. Respuestas Antes

6.1.3.1.1. Iteración 1

En la Tabla 6.75, se puede observar que para el caso de la técnica “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” y “Estructural” no coincide ya que en ambos casos la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(Before)	AppliedTec	Mean	Std. Deviation	N
Review	Review	71,4286	18,44186	7
	Structural	73,1271	17,09606	7
	Functional	79,5900	23,28237	7
Structural	Review	57,1433	18,06767	6
	Structural	68,6483	34,08573	6
	Functional	77,7767	8,86264	6
Functional	Review	41,9053	23,82119	15
	Structural	67,4593	36,32813	15
	Functional	85,2360	12,54078	15

Tabla 6.75. Estadísticos Descriptivos

En la Tabla 6.76, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$). En la Tabla 6.77, se observa que no se puede asumir esfericidad ($\text{sig} \leq 0.05$).

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.78 como en la Tabla 6.79, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con

respecto a la interacción de la técnica que los sujetos aplicaron y la técnica más fácil de aplicar, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Box's M ^a	24,067
F	1,551
df1	12
df2	1085,151
Sig.	,100

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

Tabla 6.76. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,694	8,768	2	,012	,766	,871	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.77. Test de esfericidad de Mauchly

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,460	10,236b	2,000	24,000	,001
	Wilks' Lambda	,540	10,236b	2,000	24,000	,001
	Hotelling's Trace	,853	10,236b	2,000	24,000	,001
	Roy's Largest Root	,853	10,236b	2,000	24,000	,001
AppliedTec * PT3(Before)	Pillai's Trace	,278	2,016	4,000	50,000	,106
	Wilks' Lambda	,722	2,119b	4,000	48,000	,093
	Hotelling's Trace	,384	2,210	4,000	46,000	,083
	Roy's Largest Root	,384	4,801c	2,000	25,000	,017

a. Design: Intercept + Before_EasyApply

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.78. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	6927,023	1,531	4523,463	5,110	,017
	Huynh-Feldt	6927,023	1,742	3976,911	5,110	,013
	Lower-bound	6927,023	1,000	6927,023	5,110	,033
AppliedTec * PT3(Before)	Greenhouse-Geisser	3390,434	3,063	1107,005	1,251	,305
	Huynh-Feldt	3390,434	3,484	973,250	1,251	,304
	Lower-bound	3390,434	2,000	1695,217	1,251	,304
Error(AppliedTec)	Greenhouse-Geisser	33886,502	38,284	885,138		
	Huynh-Feldt	33886,502	43,545	778,191		
	Lower-bound	33886,502	25,000	1355,460		

Tabla 6.79. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$), sin embargo, Roy's Largest Root es significativo en la interacción ($\text{sig} \leq 0.05$).

En la Tabla 6.80, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	343155,598	1	343155,598	892,088	,000
PT3(Before)	1389,015	2	694,507	1,805	,185
Error	9616,640	25	384,666		

Tabla 6.80. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias sirven para predecir. Por lo tanto, pueden garantizar que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.81, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y entre los que aplicaron “Estructural” y “Revisión” la técnica con menor media es “Revisión”. Esto se puede observar gráficamente en la Figura 6.11.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	56,826	4,407	47,749	65,903
Structural	69,745	6,598	56,156	83,334
Functional	80,868	3,127	74,428	87,307

Tabla 6.81. Medias estimadas

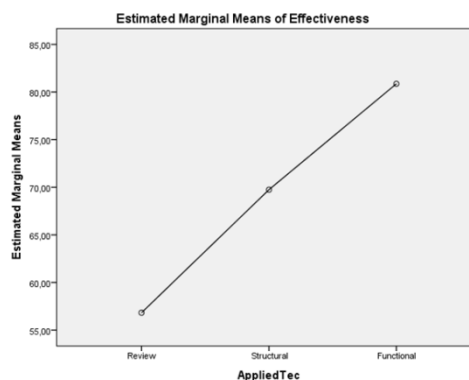


Figura 6.11. Técnica aplicada

Finalmente, en la Tabla 6.82, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias significativas entre la

“Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	-12,919	9,169	,513	-36,447	10,608
	Functional	-24,042*	5,458	,001	-38,048	-10,036
Structural	Review	12,919	9,169	,513	-10,608	36,447
	Functional	-11,123	7,490	,450	-30,343	8,097
Functional	Review	24,042*	5,458	,001	10,036	38,048
	Structural	11,123	7,490	,450	-8,097	30,343

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.82. Comparación por pares

Con respecto a la interacción de la técnica aplicada y la técnica que más fácil de aplicar, en la Tabla 6.83 se puede observar que de los que responden “Funcional” coincide con la técnica con la que han encontrado más faltas, cuando responden “Revisión” y “Estructural” encuentran más faltas en “Funcional”.

PT3(Before)	AppliedTec	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Review	Review	71,429	8,148	54,648	88,209
	Structural	73,127	12,198	48,005	98,250
	Functional	79,590	5,780	67,685	91,495
Structural	Review	57,143	8,800	39,019	75,268
	Structural	68,648	13,176	41,513	95,784
	Functional	77,777	6,243	64,918	90,635
Functional	Review	41,905	5,566	30,442	53,368
	Structural	67,459	8,333	50,297	84,621
	Functional	85,236	3,949	77,104	93,368

Tabla 6.83. Medias estimadas

En la Figura 6.12 se puede observar, que cuando los sujetos responden “Revisión” encuentran más faltas cuando aplican “Revisión”, aunque no hay mucha diferencia entre las tres técnicas. Cuando responden “Estructural” encuentran más faltas cuando aplican “Funcional”, aunque no hay mucha diferencia con “Estructural”, y encuentran menos faltas cuando aplican “Revisión”. Finalmente, cuando responden “Funcional” encuentran más faltas cuando aplican la misma y encuentran menos faltas cuando aplican “Revisión”.

Finalmente, en la Tabla 6.84, se puede comprobar que de los que dicen que la técnica que les gusta más es “Revisión”, no hay diferencias significativas entre “Revisión” y “Estructural”. Además entre “Revisión” y “Funcional” tampoco hay diferencias significativas. Finalmente tampoco hay diferencias significativas entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

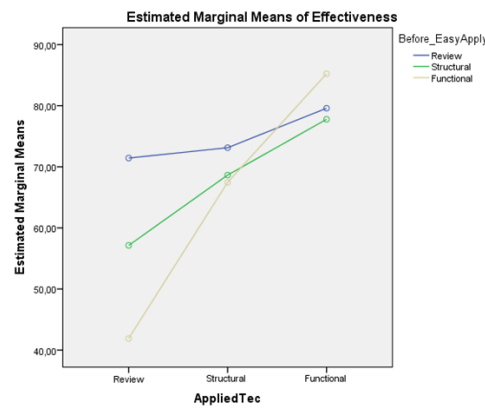


Figura 6.12 Técnica aplicada * Técnica Fácil de Aplicar (Antes)

Por otro lado de los que dicen que la técnica más efectiva es “Estructural”, no hay diferencias significativas entre “Revisión” y “Estructural”. Además entre “Revisión” y “Funcional” tampoco hay diferencias significativas. Finalmente tampoco hay diferencias significativas entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

Finalmente de los que dicen que la técnica más efectiva es “Funcional”, aplican significativamente mejor la misma que “Revisión” pero no que “Estructural”. Además no hay diferencias significativas entre “Revisión” y “Estructural” ni tampoco entre “Estructural” y “Funcional”. Por lo tanto en cierto grado coincide lo que responden los sujetos con lo que aplican.

PT3(Before)	(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Review	Review	Structural	-1,699	16,951	1,000	-45,194	41,797
		Functional	-8,161	10,091	1,000	-34,054	17,731
	Structural	Review	1,699	16,951	1,000	-41,797	45,194
		Functional	-6,463	13,847	1,000	-41,995	29,069
Structural	Functional	Review	8,161	10,091	1,000	-17,731	34,054
		Structural	6,463	13,847	1,000	-29,069	41,995
	Review	Structural	-11,505	18,309	1,000	-58,486	35,476
		Functional	-20,633	10,899	,210	-48,601	7,334
		Structural	11,505	18,309	1,000	-35,476	58,486
		Functional	-9,128	14,957	1,000	-47,508	29,251
Functional	Functional	Review	20,633	10,899	,210	-7,334	48,601
		Structural	9,128	14,957	1,000	-29,251	47,508
	Review	Structural	-25,554	11,580	,110	-55,267	4,159
		Functional	-43,331*	6,893	,000	-61,019	-25,642
		Structural	25,554	11,580	,110	-4,159	55,267
		Functional	-17,777	9,460	,216	-42,050	6,496
	Functional	Review	43,331*	6,893	,000	25,642	61,019
		Structural	17,777	9,460	,216	-6,496	42,050

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.84. Comparación por pares

6.1.3.1.2. Iteración 2

En la Tabla 6.85, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(Before)	AppliedTec	Mean	Std. Deviation	N
Review	Review	54,1667	17,25164	8
	Structural	35,4167	20,77372	8
	Functional	50,0000	15,43033	8
Structural	Review	60,0000	19,00292	5
	Structural	50,0000	16,66667	5
	Functional	33,3333	11,78511	5
Functional	Review	55,3763	22,92786	31
	Structural	50,0000	14,90712	31
	Functional	48,9246	18,72593	31

Tabla 6.85. Estadísticos Descriptivos

En la Tabla 6.86, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	24,003
F	1,527
df1	12
df2	631,986
Sig.	,110

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

Tabla 6.86. Test de igualdad de matrices de covarianzas

En la Tabla 6.87, se observa que se puede asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,959	1,682	2	,431	,960	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.87. Test de esfericidad de Mauchly

El siguiente paso es analizar los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.88, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}<0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica más fácil de aplicar, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	2393,004	2	1196,502	3,608	,031
AppliedTec * PT3(Before)	Sphericity Assumed	2069,443	4	517,361	1,560	,193
Error(AppliedTec)	Sphericity Assumed	27189,946	82	331,585		

Tabla 6.88. Test univariado de efectos intra-sujetos

En la Tabla 6.89, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	178357,774	1	178357,774	467,290	,000
PT3(Before)	552,472	2	276,236	,724	,491
Error	15649,091	41	381,685		

Tabla 6.89. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.90, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Estructural” y “Funcional”, aunque no hay mucha diferencia, la técnica con menor media es “Estructural”. Esto se puede observar gráficamente en la Figura 6.13.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	56,514	4,323	47,785	65,244
Structural	45,139	3,233	38,609	51,669
Functional	44,086	3,512	36,993	51,179

Tabla 6.90. Medias estimadas

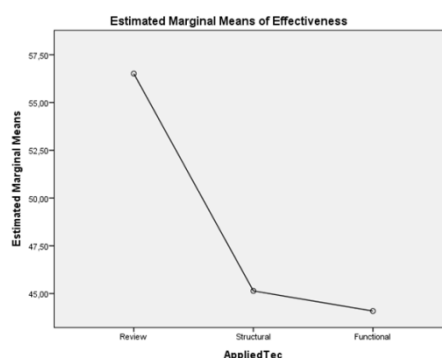


Figura 6.13. Técnica aplicada

Finalmente, en la Tabla 6.91, se puede comprobar que no hay diferencias significativas entre “Revisión” y “Estructural”. Por otro lado tampoco existen diferencias significativas entre la

“Revisión” y la técnica “Funcional”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	11,375	5,259	,109	-1,753	24,504
	Functional	12,428	5,490	,087	-1,275	26,132
Structural	Review	-11,375	5,259	,109	-24,504	1,753
	Functional	1,053	4,602	1,000	-10,434	12,540
Functional	Review	-12,428	5,490	,087	-26,132	1,275
	Structural	-1,053	4,602	1,000	-12,540	10,434

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.91. Comparación por pares

6.1.3.1.3. Iteración 3

En la Tabla 6.92, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(Before)	AppliedTec	Mean	Std. Desviation	N
Review	Review	45,2400	26,72687	7
	Structural	30,9514	29,54550	7
	Functional	45,2371	8,13413	7
Structural	Review	54,5455	21,20339	11
	Structural	43,9382	22,69561	11
	Functional	36,3627	12,51197	11
Functional	Review	55,5554	20,06471	24
	Structural	43,7504	18,91509	24
	Functional	41,6663	13,90140	24

Tabla 6.92. Estadísticos Descriptivos

En la Tabla 6.93, se puede observar que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	9,693
F	,677
df1	12
df2	1598,146
Sig.	,775

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

Tabla 6.93. Test de igualdad de matrices de covarianzas

En la Tabla 6.94, se puede observar que no se puede asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar el test multivariado y en los test univariados el valor de Huyn-Feldt.

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,798	8,576	2	,014	,832	,909	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(Before)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.94. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.95 como en la Tabla 6.96, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica más fácil de aplicar, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,155	3,496b	2,000	38,000	,040
	Wilks' Lambda	,845	3,496b	2,000	38,000	,040
	Hotelling's Trace	,184	3,496b	2,000	38,000	,040
	Roy's Largest Root	,184	3,496b	2,000	38,000	,040
AppliedTec * PT3(Before)	Pillai's Trace	,137	1,434	4,000	78,000	,231
	Wilks' Lambda	,863	1,452b	4,000	76,000	,225
	Hotelling's Trace	,159	1,468	4,000	74,000	,221
	Roy's Largest Root	,159	3,094c	2,000	39,000	,057

a. Design: Intercept + Before_EasyApply

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.95. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	2901,060	1,664	1743,592	4,089	,028
	Huynh-Feldt	2901,060	1,818	1595,644	4,089	,024
	Lower-bound	2901,060	1,000	2901,060	4,089	,050
AppliedTec * PT3(Before)	Greenhouse-Geisser	1229,754	3,328	369,553	,867	,472
	Huynh-Feldt	1229,754	3,636	338,195	,867	,480
	Lower-bound	1229,754	2,000	614,877	,867	,428
Error(AppliedTec)	Greenhouse-Geisser	27668,513	64,890	426,392		
	Huynh-Feldt	27668,513	70,906	390,212		
	Lower-bound	27668,513	39,000	709,449		

Tabla 6.96. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$).

En la Tabla 6.97, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	190978,388	1	190978,388	483,722	,000
PT3(Before)	697,910	2	348,955	,884	,421
Error	15397,596	39	394,810		

Tabla 6.97. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.98, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Estructural” y “Funcional”, aunque no hay mucha diferencia, la técnica con menor media es “Funcional”. Esto se puede observar gráficamente en la Figura 6.14.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	51,780	3,763	44,170	59,391
Structural	39,547	3,822	31,816	47,278
Functional	41,089	2,242	36,553	45,624

Tabla 6.98. Medias estimadas

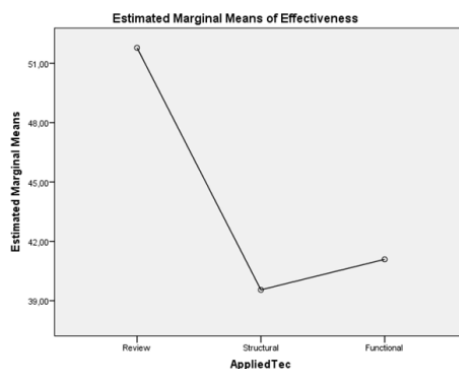


Figura 6.14. Técnica aplicada

Finalmente, en la Tabla 6.99, se puede comprobar que “Revisión” es significativamente mejor que “Funcional”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	12,234	5,610	,106	-1,800	26,267
	Functional	10,692*	4,077	,037	,493	20,890
Structural	Review	-12,234	5,610	,106	-26,267	1,800
	Functional	-1,542	4,129	1,000	-11,871	8,787
Functional	Review	-10,692*	4,077	,037	-20,890	-,493
	Structural	1,542	4,129	1,000	-8,787	11,871

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.99. Comparación por pares

6.1.3.1.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.100, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas que en “Revisión”. Con respecto a la interacción de la técnica aplicada y la técnica más fácil de aplicar, cuando los sujetos responden “Revisión” y “Estructural” encuentran el mismo número de faltas con las tres técnicas y cuando responden “Funcional” encuentran más faltas en la misma que en “Revisión” pero no que en “Estructural”. Además no influyen las preferencias que se tienen sobre las técnicas.

En el caso de la iteración 2, la técnica aplicada sale significativa, sin embargo, no hay ninguna técnica que se haya aplicado mejor que las otras. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Finalmente, en la iteración 3 cuando los sujetos aplican la técnica “Revisión” encuentran más faltas que en “Funcional”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Debido a que en dos iteraciones hay técnicas que detectan más faltas y estas no son la misma, se podría decir, que realmente no es la técnica la que determina este comportamiento. Es decir, al parecer existen otros factores, como las cualidades de los sujetos o la preparación previa al estudio empírico, que influyen al aplicar las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,001	F>R	,185	N/A	,106	R R=E=F
		R=E				E R=E=F
		E=F				F F>R
						R=E
Iteración 2	,031	R=E=F	,491	N/A	,193	N/A
Iteración 3	,040	R>F	,421	N/A	,231	N/A
		R=E				
		E=F				

Tabla 6.100. Resumen

6.1.3.2. Respuestas Después

6.1.3.2.1. Iteración 1

En la Tabla 6.101, se puede observar que para el caso de la técnica “Estructural” y “Funcional” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Revisión” no coincide ya que en ambos casos la que mejor aplican es “Funcional”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	75,0000	24,39555	4
	Structural	81,5450	6,83603	4
	Functional	84,5225	13,67583	4
Structural	Review	47,6200	26,59865	6
	Structural	77,7750	11,43422	6
	Functional	76,9817	23,21501	6
Functional	Review	49,6242	21,48882	19
	Structural	64,6605	36,68992	19
	Functional	84,3342	12,77986	19

Tabla 6.101. Estadísticos Descriptivos

En la Tabla 6.102, se observa que no se puede asumir que las matrices de covarianzas son iguales ($\text{sig} \leq 0.05$).

Box's M ^a	41,764
F	2,328
df1	12
df2	364,841
Sig.	,007

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

Tabla 6.102. Test de igualdad de matrices de covarianzas

Aunque no se puede evaluar el test de Mauchly, en la Tabla 6.103, se puede observar que tampoco hubiese sido posible asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados el valor de Huynh-Feldt.

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,760	6,845	2	,033	,807	,918	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.103. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huynh-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.104

como en la Tabla 6.105, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,349	6,699b	2,000	25,000	,005
	Wilks' Lambda	,651	6,699b	2,000	25,000	,005
	Hotelling's Trace	,536	6,699b	2,000	25,000	,005
	Roy's Largest Root	,536	6,699b	2,000	25,000	,005
AppliedTec * PT3(After)	Pillai's Trace	,147	1,032	4,000	52,000	,400
	Wilks' Lambda	,856	1,008b	4,000	50,000	,412
	Hotelling's Trace	,164	,983	4,000	48,000	,426
	Roy's Largest Root	,134	1,740c	2,000	26,000	,195

a. Design: Intercept + After_EasyApply

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.104. Test multivariado de efectos intra-sujetos

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	6087,467	1,614	3772,776	4,475	,024
	Huynh-Feldt	6087,467	1,837	3314,234	4,475	,019
	Lower-bound	6087,467	1,000	6087,467	4,475	,044
AppliedTec * PT3(After)	Greenhouse-Geisser	2012,870	3,227	623,749	,740	,543
	Huynh-Feldt	2012,870	3,674	547,939	,740	,559
	Lower-bound	2012,870	2,000	1006,435	,740	,487
Error(AppliedTec)	Greenhouse-Geisser	35366,749	41,952	843,036		
	Huynh-Feldt	35366,749	47,756	740,574		
	Lower-bound	35366,749	26,000	1360,260		

Tabla 6.105. Test univariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$).

En la Tabla 6.106, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	292809,563	1	292809,563	802,372	,000
PT3(After)	2005,519	2	1002,759	2,748	,083
Error	9488,184	26	364,930		

Tabla 6.106. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas

- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.107, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Funcional” y entre los que aplicaron “Estructural” y “Revisión” la técnica con menor media es “Revisión”. Esto se puede observar gráficamente en la Figura 6.15.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	57,415	5,229	46,666	68,164
Structural	74,660	7,084	60,098	89,222
Functional	81,946	3,525	74,700	89,192

Tabla 6.107. Medias estimadas

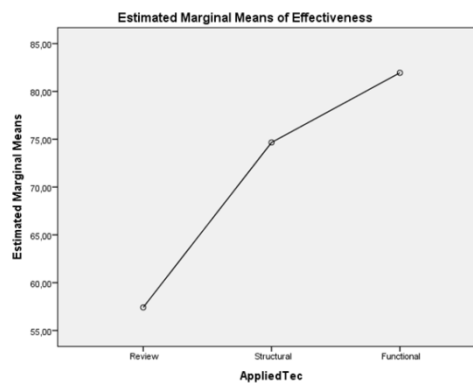


Figura 6.15. Técnica aplicada

Finalmente, en la Tabla 6.108, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	-17,245	10,191	,308	-43,324	8,833
	Functional	-24,531*	6,705	,003	-41,688	-7,375
Structural	Review	17,245	10,191	,308	-8,833	43,324
	Functional	-7,286	7,999	1,000	-27,754	13,182
Functional	Review	24,531*	6,705	,003	7,375	41,688
	Structural	7,286	7,999	1,000	-13,182	27,754

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.108. Comparación por pares

6.1.3.2.2. Iteración 2

En la Tabla 6.109, se puede observar que para el caso de la técnica “Revisión” y “Estructural” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Funcional” no coincide ya que la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	59,2593	20,60055	9
	Structural	50,0000	14,43376	9
	Functional	49,9996	14,43424	9
Structural	Review	50,0000	28,86751	3
	Structural	50,0000	16,66667	3
	Functional	22,2222	25,45875	3
Functional	Review	50,6410	23,79578	26
	Structural	44,2308	18,82215	26
	Functional	48,7179	16,27751	26

Tabla 6.109. Estadísticos Descriptivos

En la Tabla 6.110, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	1,931
F	,273
df1	6
df2	1359,451
Sig.	,950

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

Tabla 6.110. Test de igualdad de matrices de covarianzas

En la Tabla 6.111, se puede observar que se puede asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,949	1,784	2	,410	,951	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.111. Test de esfericidad de Mauchly

El siguiente paso es analizar los test univariados de los efectos intra-sujetos. Con respecto a la técnica aplicada, en la Tabla 6.112, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig}>0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Sphericity Assumed	1591,699	2	795,849	2,106	,129
AppliedTec * PT3(After)	Sphericity Assumed	1650,516	4	412,629	1,092	,367
Error(AppliedTec)	Sphericity Assumed	26453,910	70	377,913		

Tabla 6.112. Test univariado de efectos intra-sujetos

En la Tabla 6.113, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	124720,779	1	124720,779	325,039	,000
PT3(After)	1141,237	2	570,618	1,487	,240
Error	13429,866	35	383,710		

Tabla 6.113. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Todas las técnicas detectan el mismo número de faltas
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

6.1.3.2.3. Iteración 3

En la Tabla 6.114, se puede observar que para el caso de la técnica “Revisión” coincide la técnica elegida con la técnica que han aplicado mejor, en el caso de “Estructural” y “Funcional” no coincide ya que en ambos casos la que mejor aplican es “Revisión”. A continuación, se va a analizar si esta coincidencia es significativa.

PT3(After)	AppliedTec	Mean	Std. Deviation	N
Review	Review	61,6680	17,65695	10
	Structural	33,3320	29,39556	10
	Functional	41,6660	11,78511	10
Structural	Review	45,8325	8,33500	4
	Structural	41,6675	21,51700	4
	Functional	33,3325	13,60692	4
Functional	Review	52,0834	22,30112	32
	Structural	43,7500	22,30001	32
	Functional	40,6244	15,22744	32

Tabla 6.114. Estadísticos Descriptivos

En la Tabla 6.115, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

En la Tabla 6.116, se observa que no se puede asumir esfericidad ($\text{sig} \leq 0.05$), por lo tanto, el siguiente paso es analizar los test multivariados y en los test univariados el valor de Huyn-Feldt.

Box's M ^a	10,227
F	,612
df1	12
df2	325,813
Sig.	,832

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

Tabla 6.115. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedTec	,852	6,746	2	,034	,871	,947	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PT3(After)

Within Subjects Design: AppliedTec

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.116. Test de esfericidad de Mauchly

Dado que no se puede asumir esfericidad, se tomará el valor de los test multivariados y en los test univariados el valor de Huyn-Feldt. Con respecto a la técnica aplicada, tanto en la Tabla 6.117 como en la Tabla 6.118, se puede observar que para la técnica aplicada el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción de la técnica que los sujetos aplicaron y la técnica que más les gusta, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Effect ^a		Value	F	Hypothesis df	Error df	Sig.
AppliedTec	Pillai's Trace	,181	4,652b	2,000	42,000	,015
	Wilks' Lambda	,819	4,652b	2,000	42,000	,015
	Hotelling's Trace	,222	4,652b	2,000	42,000	,015
	Roy's Largest Root	,222	4,652b	2,000	42,000	,015
AppliedTec * PT3(After)	Pillai's Trace	,072	,806	4,000	86,000	,525
	Wilks' Lambda	,928	,801b	4,000	84,000	,528
	Hotelling's Trace	,078	,796	4,000	82,000	,531
	Roy's Largest Root	,076	1,631c	2,000	43,000	,208

a. Design: Intercept + After_EasyApply

Within Subjects Design: AppliedTec

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Tabla 6.117. Test multivariado de efectos intra-sujetos

Todos los estadístico de los test multivariados coinciden con los estadísticos de los test univariados, para la técnica aplicada ($\text{sig} \leq 0.05$) y para la interacción de la técnica aplicada con la elegida ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedTec	Greenhouse-Geisser	3156,111	1,742	1812,214	4,206	,023
	Huynh-Feldt	3156,111	1,893	1667,032	4,206	,020
	Lower-bound	3156,111	1,000	3156,111	4,206	,046
AppliedTec * PT3(After)	Greenhouse-Geisser	1710,012	3,483	490,938	1,140	,342
	Huynh-Feldt	1710,012	3,787	451,607	1,140	,343
	Lower-bound	1710,012	2,000	855,006	1,140	,329
Error(AppliedTec)	Greenhouse-Geisser	32264,121	74,888	430,833		
	Huynh-Feldt	32264,121	81,410	396,317		
	Lower-bound	32264,121	43,000	750,328		

Tabla 6.118. Test univariado de efectos intra-sujetos

En la Tabla 6.119, se observa que las opiniones de la técnica más fácil de aplicar no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	135695,367	1	135695,367	295,540	,000
After_EasyApply	299,225	2	149,613	,326	,724
Error	19743,152	43	459,143		

Tabla 6.119. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay técnicas que detectan más faltas que otras
- Las preferencias de los sujetos sobre una técnica no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que una técnica se vaya a aplicar mejor que otra.

Con respecto a la técnica aplicada, en la Tabla 6.120, se puede observar que la media más alta la tienen los sujetos que aplicaron la técnica “Revisión” y entre los que aplicaron “Estructural” y “Funcional”, aunque no hay mucha diferencia, la técnica con menor media es “Funcional”. Esto se puede observar gráficamente en la Figura 6.16.

AppliedTec	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Review	53,195	4,261	44,601	61,788
Structural	39,583	4,921	29,659	49,507
Functional	38,541	2,977	32,538	44,544

Tabla 6.120. Medias estimadas

Finalmente, en la Tabla 6.121, se puede comprobar que la técnica “Funcional” es significativamente mejor que la “Revisión”. Por otro lado no existen diferencias significativas entre la “Revisión” y la técnica “Estructural”. Finalmente también se puede comprobar que tampoco existen diferencias significativas entre la técnica “Estructural” y la técnica “Funcional”.

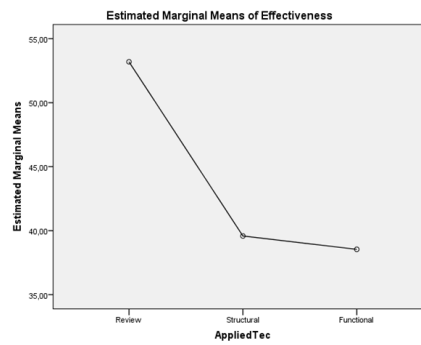


Figura 6.16. Técnica aplicada

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Review	Structural	13,611	6,593	,135	-2,814	30,037
	Functional	14,654*	4,772	,011	2,764	26,543
Structural	Review	-13,611	6,593	,135	-30,037	2,814
	Functional	1,042	5,395	1,000	-12,399	14,483
Functional	Review	-14,654*	4,772	,011	-26,543	-2,764
	Structural	-1,042	5,395	1,000	-14,483	12,399

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.121. Comparación por pares

6.1.3.2.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.122, se observa que en la iteración 1, cuando los sujetos aplican la técnica “Funcional” encuentran más faltas que en “Revisión”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Finalmente, en la iteración 3 cuando los sujetos aplican la técnica “Revisión” encuentran más faltas que en “Funcional”. Además no influyen ni las preferencias que se tienen sobre las técnicas, ni la interacción de técnica elegida con la técnica aplicada.

Debido a que en dos iteraciones hay técnicas que detectan más faltas y estas no son la misma, se podría decir, que realmente no es la técnica la que determina este comportamiento. Es decir, al parecer existen otros factores, que influyen al aplicar las técnicas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,005	F>R	,083	N/A	,400	N/A
		R=E				
		E=F				
Iteración 2	,129	N/A	,240	N/A	,367	N/A
Iteración 3	,015	R>F	,724	N/A	,525	N/A
		R=E				
		E=F				

Tabla 6.122. Resumen

6.2. Preferencias sobre los programas

6.2.1. ¿Qué programa es más fácil?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa que los sujetos consideran que es el más fácil. Esta pregunta aparece en las tres iteraciones; además únicamente se preguntó a los sujetos después del estudio empírico.

6.2.1.1. Iteración 1

En la Tabla 6.123, se puede observar que para el caso del programa "Cmdline" y "Ntree" coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de "Nametbl" no coincide ya que el programa en el que han encontrado más faltas es "Cmdline". A continuación, se va a analizar si esta coincidencia es significativa.

PP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	72,0225	32,61447	8
	Nametbl	57,1425	30,54061	8
	Ntree	64,2838	30,54228	8
Nametbl	Cmdline	84,9200	14,35121	3
	Nametbl	52,3800	45,92089	3
	Ntree	71,4300	28,57000	3
Ntree	Cmdline	60,7130	28,33224	10
	Nametbl	67,1410	17,87867	10
	Ntree	87,1430	14,20547	10

Tabla 6.123. Estadísticos Descriptivos

En la Tabla 6.124, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	7,666
F	1,008
df1	6
df2	1582,708
Sig.	,418

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

Tabla 6.124. Test de igualdad de matrices de covarianzas

En la Tabla 6.125, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.126, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,979	,356	2	,837	,980	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.125. Test de esfericidad de Mauchly

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	2293,262	2	1146,631	1,344	,274
AppliedProg * PP1	Sphericity Assumed	3936,947	4	984,237	1,153	,347
Error(AppliedProg)	Sphericity Assumed	30721,164	36	853,366		

Tabla 6.126. Test univariado de efectos intra-sujetos

En la Tabla 6.127, se observa que las opiniones del programa más fácil no son significativas (sig>0.05).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	227406,511	1	227406,511	514,044	,000
PP1	697,283	2	348,641	,788	,470
Error	7962,964	18	442,387		

Tabla 6.127. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- En todos los programas se detectan el mismo número de faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

6.2.1.2. Iteración 2

En la Tabla 6.128, se puede observar que para el caso del programa "Nametbl" y "Ntree" coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de "Cmdline" no coincide ya que el programa en el que han encontrado más faltas es "Nametbl". A continuación, se va a analizar si esta coincidencia es significativa.

PP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	39,9990	22,49752	10
	Nametbl	58,3340	16,19861	10
	Ntree	48,3320	16,57478	10
Nametbl	Cmdline	43,7500	20,97212	16
	Nametbl	54,1675	12,91253	16
	Ntree	52,0838	22,66924	16
Ntree	Cmdline	33,3333	18,80227	12
	Nametbl	48,6108	15,00861	12
	Ntree	55,5558	20,51517	12

Tabla 6.128. Estadísticos Descriptivos

En la Tabla 6.129, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	7,887
F	,570
df1	12
df2	4265,317
Sig.	,868

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

Tabla 6.129. Test de igualdad de matrices de covarianzas

En la Tabla 6.130, se observa que también es posible asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,960	1,399	2	,497	,961	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.130. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.131, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	4715,176	2	2357,588	7,192	,001
AppliedProg * PP1	Sphericity Assumed	1193,163	4	298,291	,910	,463
Error(AppliedProg)	Sphericity Assumed	22945,534	70	327,793		

Tabla 6.131. Test univariado de efectos intra-sujetos

En la Tabla 6.132, se observa que las opiniones del programa más fácil no son significativas ($\text{sig}>0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	255593,671	1	255593,671	627,913	,000
PP1	367,251	2	183,625	,451	,641
Error	14246,852	35	407,053		

Tabla 6.132. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas

- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.133, se puede observar que la media más alta tienen los sujetos que aplicaron el programa “Nametbl” y la más baja los que aplicaron el programa “Cmdline”. Esto se puede comprobar gráficamente en la Figura 6.17.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	39,027	3,426	32,072	45,983
Nametbl	53,704	2,393	48,845	58,563
Ntree	51,991	3,400	45,089	58,892

Tabla 6.133. Medias estimadas

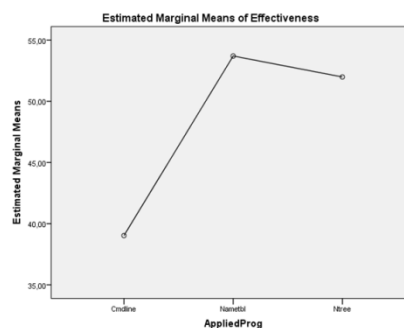


Figura 6.17. Programa Aplicado

Finalmente, en la Tabla 6.134, se puede comprobar que el programa “Nametbl” es significativamente mejor que “Cmdline” y también “Ntree” es significativamente mejor que “Cmdline”. Finalmente no hay diferencias significativas entre el programa “Nametbl” y “Ntree”

(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-14,677*	3,994	,002	-24,719	-4,634
	Ntree	-12,963*	4,637	,025	-24,622	-1,304
Nametbl	Cmdline	14,677*	3,994	,002	4,634	24,719
	Ntree	1,714	4,034	1,000	-8,431	11,858
Ntree	Cmdline	12,963*	4,637	,025	1,304	24,622
	Nametbl	-1,714	4,034	1,000	-11,858	8,431

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.134. Comparación por pares

6.2.1.3. Iteración 3

En la Tabla 6.135, se puede observar que para el caso del programa” Nametbl” y “Ntree” coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de “Cmdline” no coincide ya que el programa en el que han encontrado más faltas es “Nametbl”. A continuación, se va a analizar si esta coincidencia es significativa.

PP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	41,6665	17,52267	20
	Nametbl	57,5000	19,84929	20
	Ntree	47,5000	21,81461	20
Nametbl	Cmdline	40,9073	17,26168	11
	Nametbl	53,0309	19,46294	11
	Ntree	40,9082	22,80778	11
Ntree	Cmdline	33,3336	18,48954	14
	Nametbl	48,8093	19,02009	14
	Ntree	36,9043	22,81424	14

Tabla 6.135. Estadísticos Descriptivos

En la Tabla 6.136, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	2,184
F	,161
df1	12
df2	5223,411
Sig.	,999

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

Tabla 6.136. Test de igualdad de matrices de covarianzas

En la Tabla 6.137, se puede observar que también es posible asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,972	1,164	2	,559	,973	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.137. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.138, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	4917,883	2	2458,941	6,432	,003
AppliedProg * PP1	Sphericity Assumed	131,031	4	32,758	,086	,987
Error(AppliedProg)	Sphericity Assumed	32112,640	84	382,293		

Tabla 6.138. Test univariado de efectos intra-sujetos

En la Tabla 6.139, se observa que las opiniones del programa más fácil no son significativas ($\text{sig}>0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	251876,056	1	251876,056	583,434	,000
PP1	2094,595	2	1047,297	2,426	,101
Error	18131,932	42	431,713		

Tabla 6.139. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.140, se observa que la media más alta y más baja está en “Nametbl” y “Cmdline”, respectivamente. Esto se puede comprobar en la Figura 6.18.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	38,636	2,729	33,128	44,143
Nametbl	53,113	2,996	47,068	59,159
Ntree	41,771	3,435	34,838	48,704

Tabla 6.140. Medias estimadas

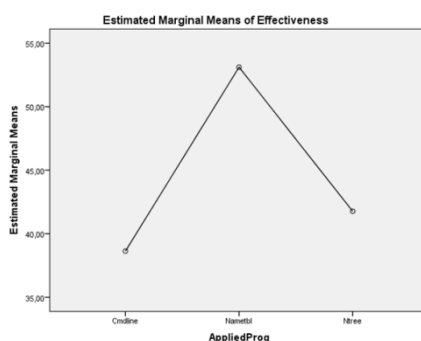


Figura 6.18. Programa Aplicado

Finalmente, en la Tabla 6.141, se puede comprobar que el programa “Nametbl” es significativamente mejor que “Cmdline”. Por otro lado, no hay diferencias significativas entre “Cmdline” y “Ntree”. Finalmente no hay diferencias significativas entre “Nametbl” y “Ntree”

(I) AppliedProg	(J) AppliedProg	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-14,478*	4,182	,004	-24,905	-4,050
	Ntree	-3,135	3,968	1,000	-13,030	6,760
Nametbl	Cmdline	14,478*	4,182	,004	4,050	24,905
	Ntree	11,343	4,570	,051	-,053	22,739
Ntree	Cmdline	3,135	3,968	1,000	-6,760	13,030
	Nametbl	-11,343	4,570	,051	-22,739	,053

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.141. Comparación por pares

6.2.1.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.142, se observa que en la iteración 2, cuando los sujetos aplican el programa “Cmdline” encuentran menos faltas en el mismo que en “Nametbl” y “Ntree”. Además no influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Finalmente, en la iteración 3 cuando los sujetos aplican el programa “Nametbl” encuentran más faltas en el mismo que en “Cmdline” pero no que en “Ntree”. Además no influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Debido a que en dos iteraciones hay programas en los que se detectan más faltas, se podría decir, que realmente en ciertos programas es más fácil encontrar las faltas que en otros. Sin embargo, dado que en la iteración 1, no hay ningún programa mejor, es probable que existan otros factores que influyen al aplicar los programas.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,247	N/A	,470	N/A	,347	N/A
Iteración 2	,001	(na=nt)>cm	,641	N/A	,463	N/A
Iteración 3	,009	na>cm	,304	N/A	,606	N/A
		cm=nt				
		na=nt				

Tabla 6.142. Resumen

6.2.2. ¿Qué programa tiene las faltas más difíciles?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa que los sujetos consideran que tiene las faltas más difíciles. Esta pregunta aparece en las tres iteraciones; además únicamente se preguntó a los sujetos después del estudio empírico.

6.2.2.1. Iteración 1

En la Tabla 6.143, se puede observar que cuando responden “Cmdline” coincide con el programa en el que han encontrado menos faltas, en el caso de “Nametbl” y “Ntree” no coincide ya que el programa en el que han encontrado menos faltas es “Ntree” y “Nametbl”, respectivamente. A continuación, se va a analizar si esta coincidencia es significativa.

En la Tabla 6.144, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$). En la Tabla 6.145, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.146, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica es el mismo ($\text{sig} > 0.05$). Por otro lado, con respecto a la interacción

del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

PP2	AppliedProg	Mean	Std. Desviation	N
Cmdline	Cmdline	53,5700	37,56971	4
	Nametbl	60,7150	21,42556	4
	Ntree	75,0000	24,39555	4
Nametbl	Cmdline	83,3300	.	1
	Nametbl	57,1400	.	1
	Ntree	,0000	.	1
Ntree	Cmdline	70,6333	23,55474	6
	Nametbl	47,6183	33,39934	6
	Ntree	83,3317	16,70109	6

Tabla 6.143. Estadísticos Descriptivos

Box's M ^a	22,682
F	2,023
df1	6
df2	265,544
Sig.	,063

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

Tabla 6.144. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,986	,095	2	,954	,987	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.145. Test de esfericidad de Mauchly

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	997,847	2	498,924	,553	,586
AppliedProg * PP2	6380,309	4	1595,077	1,769	,184
Error(AppliedProg)	14425,452	16	901,591		

Tabla 6.146. Test univariado de efectos intra-sujetos

En la Tabla 6.147, se observa que las opiniones del programa más fácil no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	66428,335	1	66428,335	195,526	,000
PP2	1077,898	2	538,949	1,586	,263
Error	2717,929	8	339,741		

Tabla 6.147. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- En todos los programas se detectan el mismo número de faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

6.2.2.2. Iteración 2

En la Tabla 6.148, se puede observar que cuando responden “Cmdline” coincide con el programa en el que han encontrado menos faltas, en el caso de “Nametbl” y “Ntree” no coincide ya que el programa en el que han encontrado menos faltas también es “Cmdline”. A continuación, se va a analizar si esta coincidencia es significativa.

PP2	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	32,0523	20,92911	13
	Nametbl	51,2823	15,90441	13
	Ntree	52,5638	20,23723	13
Nametbl	Cmdline	43,9391	20,10010	11
	Nametbl	54,5464	15,07858	11
	Ntree	56,0609	21,43810	11
Ntree	Cmdline	42,8557	20,37592	14
	Nametbl	54,7621	13,75816	14
	Ntree	48,8093	20,11194	14

Tabla 6.148. Estadísticos Descriptivos

En la Tabla 6.149, se puede observar que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	8,638
F	,628
df1	12
df2	5339,821
Sig.	,820

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

Tabla 6.149. Test de igualdad de matrices de covarianzas

En la Tabla 6.150, se puede observar que también es posible asumir esfericidad ($\text{sig}>0.05$).

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.151, se puede observar que el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,963	1,277	2	,528	,964	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.150. Test de esfericidad de Mauchly

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	4515,928	2	2257,964	6,778	,002
AppliedProg * PP2	Sphericity Assumed	818,165	4	204,541	,614	,654
Error(AppliedProg)	Sphericity Assumed	23320,532	70	333,150		

Tabla 6.151. Test univariado de efectos intra-sujetos

En la Tabla 6.152, se observa que las opiniones del programa más fácil no son significativas (sig>0.05).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	265898,363	1	265898,363	668,957	,000
PP2	702,234	2	351,117	,883	,422
Error	13911,869	35	397,482		

Tabla 6.152. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.153, se puede observar que la media más baja tienen los sujetos que aplicaron el programa “Ntree” y la más alta los que aplicaron el programa “Cmdline”. Esto se puede comprobar gráficamente en la Figura 6.19.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	39,616	3,341	32,834	46,398
Nametbl	53,530	2,429	48,598	58,462
Ntree	52,478	3,349	45,679	59,277

Tabla 6.153. Medias estimadas

Finalmente, en la Tabla 6.154, se puede comprobar que el programa “Cmdline” es significativamente peor que la “Nametbl” y “Ntree”. Finalmente también se puede comprobar que no existen diferencias significativas entre el programa “Nametbl” y “Ntree”.

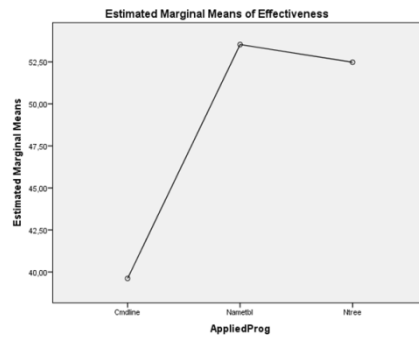


Figura 6.19. Programa Aplicado

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-13,915*	3,927	,003	-23,790	-4,039
	Ntree	-12,862*	4,587	,025	-24,395	-1,329
Nametbl	Cmdline	13,915*	3,927	,003	4,039	23,790
	Ntree	1,052	4,084	1,000	-9,217	11,322
Ntree	Cmdline	12,862*	4,587	,025	1,329	24,395
	Nametbl	-1,052	4,084	1,000	-11,322	9,217

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.154. Comparación por pares

6.2.2.3. Iteración 3

En la Tabla 6.155, se puede observar que cuando responden “Cmdline” y “Nametbl” coincide con el programa en el que han encontrado menos faltas, en el caso de “Ntree” no coincide ya que el programa en el que han encontrado menos faltas es “Cmdline”. A continuación, se va a analizar si esta coincidencia es significativa

PP2	AppliedProg	Mean	Std. Desviation	N
Cmdline	Cmdline	30,3918	20,61016	17
	Nametbl	52,9412	15,85207	17
	Ntree	41,1765	25,08184	17
Nametbl	Cmdline	45,5560	16,01980	15
	Nametbl	44,4447	22,41982	15
	Ntree	46,6660	20,11976	15
Ntree	Cmdline	39,2843	15,48086	14
	Nametbl	65,4764	13,81115	14
	Ntree	40,4757	21,39760	14

Tabla 6.155. Estadísticos Descriptivos

En la Tabla 6.156, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$). En la Tabla 6.157, se observa que también es posible asumir esfericidad ($\text{sig} > 0.05$).

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.158, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$).

Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia es significativa ($\text{sig} \leq 0.05$).

Box's M ^a	11,820
F	,884
df1	12
df2	8425,142
Sig.	,563

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

Tabla 6.156. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,992	,341	2	,843	,992	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP2

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.157. Test de esfericidad de Mauchly

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	6150,047	2	3075,023	8,960	,000
AppliedProg * PP2	Sphericity Assumed	4314,731	4	1078,683	3,143	,018
Error(AppliedProg)	Sphericity Assumed	29514,748	86	343,195		

Tabla 6.158. Test univariado de efectos intra-sujetos

En la Tabla 6.159, se observa que las opiniones del programa más fácil no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	279592,660	1	279592,660	625,763	,000
PP2	1125,785	2	562,892	1,260	,294
Error	19212,520	43	446,803		

Tabla 6.159. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias sirven para predecir. Por lo tanto, pueden garantizar que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.160, se puede observar que la media más baja tienen los sujetos que aplicaron el programa "Cmdline" y la más alta los que aplicaron el programa "Nametbl". Esto se puede comprobar gráficamente en la Figura 6.20.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	38,411	2,621	33,124	43,697
Nametbl	54,287	2,625	48,995	59,580
Ntree	42,773	3,322	36,074	49,472

Tabla 6.160. Medias estimadas

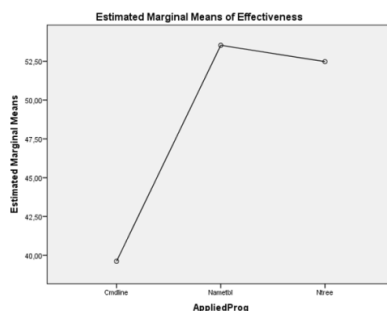


Figura 6.20. Programa Aplicado

Finalmente, en la Tabla 6.161, se puede comprobar que el programa “Cmdline” es significativamente peor que “Nametbl”. Por otro lado, el programa “Ntree” también es significativamente peor que “Nametbl”. Finalmente también se puede comprobar que no existen diferencias significativas entre el programa “Cmdline” y “Ntree”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-15,877*	3,727	,000	-25,160	-6,593
	Ntree	-4,362	3,866	,796	-13,993	5,269
Nametbl	Cmdline	15,877*	3,727	,000	6,593	25,160
	Ntree	11,515*	4,028	,020	1,481	21,548
Ntree	Cmdline	4,362	3,866	,796	-5,269	13,993
	Nametbl	-11,515*	4,028	,020	-21,548	-1,481

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.161. Comparación por pares

Con respecto a la interacción del programa elegido con el programa aplicado, en la Tabla 6.162, se puede observar que de los que respondieron que el programa con faltas más difíciles era “Cmdline” coincide con el programa en el que encontraron menos faltas. De los que respondieron que el programa con faltas más difíciles era “Nametbl” coincide con el programa en el que encontraron menos faltas, aunque no hay mucha diferencia entre los tres. Finalmente, de los que respondieron que el programa con faltas más difíciles era “Ntree” encontraron menos faltas en “Cmdline”.

En la Figura 6.21, se puede observar que los que respondieron “Cmdline” lo programas en los que encontraron más faltas son “Ntree” y “Nametbl”, puesto que entre los dos no hay mucha diferencia. En el caso de los que responden “Nametbl” encuentran más faltas en “Nametbl” y “Ntree”, entre estos dos no hay mucha diferencia. Finalmente los que eligen “Ntree” encuentran más faltas en “Nametbl”, aunque no hay mucha diferencia entre los tres programas.

PP2	AppliedProg	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Cmdline	Cmdline	30,392	4,298	21,724	39,060
	Nametbl	52,941	4,303	44,263	61,620
	Ntree	41,176	5,447	30,192	52,160
Nametbl	Cmdline	45,556	4,576	36,328	54,784
	Nametbl	44,445	4,581	35,206	53,684
	Ntree	46,666	5,798	34,973	58,359
Ntree	Cmdline	39,284	4,736	29,732	48,836
	Nametbl	65,476	4,742	55,913	75,040
	Ntree	40,476	6,002	28,372	52,579

Tabla 6.162. Medias estimadas

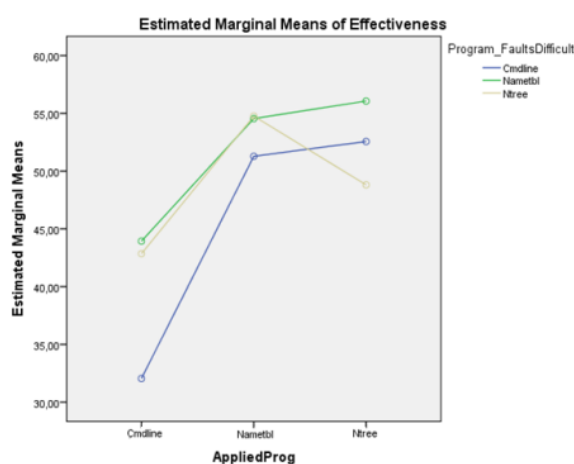


Figura 6.21. Programa con faltas más difíciles * Programa Aplicado

Finalmente, en la Tabla 6.163, se puede comprobar que de los que dicen que el programa con faltas más difíciles es “Cmdline”, “Cmdline” es significativamente peor que “Nametbl”. Por otro lado, no hay diferencias significativas entre “Cmdline” y “Ntree”. Finalmente, tampoco hay diferencias significativas entre “Nametbl” y “Ntree”. Por lo tanto parece que coincide lo que responden los sujetos con lo que sucede en la realidad.

Por otro lado de los que dicen que el programa con faltas más difíciles es “Nametbl”, no hay diferencias significativas con “Cmdline”, entre “Nametbl” y “Ntree” tampoco hay diferencias significativas. Finalmente entre “Nametbl” y “Ntree” tampoco hay diferencias significativas. Por lo tanto parece que no coincide lo que responden los sujetos con lo que sucede en realidad.

Finalmente de los que dicen que el programa con faltas más difíciles es “Ntree”, éste es significativamente peor que “Nametbl”, por otro lado, el programa “Cmdline” es significativamente peor que “Nametbl”. Finalmente entre “Cmdline” y “Ntree” no hay diferencias significativas. Por lo tanto parece que coincide lo que responden los sujetos con lo que sucede en realidad.

PP2	(I) Applied Prog	(J) Applied Prog	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
						Lower Bound	Upper Bound
Cmdline	Cmdline	Nametbl	-22,549*	6,110	,002	-37,771	-7,327
		Ntree	-10,785	6,339	,288	-26,577	5,007
	Nametbl	Cmdline	22,549*	6,110	,002	7,327	37,771
		Ntree	11,765	6,604	,246	-4,687	28,217
	Ntree	Cmdline	10,785	6,339	,288	-5,007	26,577
		Nametbl	-11,765	6,604	,246	-28,217	4,687
Nametbl	Cmdline	Nametbl	1,111	6,505	1,000	-15,094	17,316
		Ntree	-1,110	6,748	1,000	-17,922	15,702
	Nametbl	Cmdline	-1,111	6,505	1,000	-17,316	15,094
		Ntree	-2,221	7,030	1,000	-19,736	15,293
	Ntree	Cmdline	1,110	6,748	1,000	-15,702	17,922
		Nametbl	2,221	7,030	1,000	-15,293	19,736
Ntree	Cmdline	Nametbl	-26,192*	6,733	,001	-42,966	-9,418
		Ntree	-1,191	6,985	1,000	-18,593	16,211
	Nametbl	Cmdline	26,192*	6,733	,001	9,418	42,966
		Ntree	25,001*	7,277	,004	6,872	43,130
	Ntree	Cmdline	1,191	6,985	1,000	-16,211	18,593
		Nametbl	-25,001*	7,277	,004	-43,130	-6,872

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.163. Comparación por pares

6.2.2.4. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.164, se observa que en la iteración 2, cuando los sujetos aplican el programa “Cmdline” encuentran menos faltas en el mismo que en “Nametbl” y “Ntree”. Además no influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Finalmente, en la iteración 3 cuando los sujetos aplican el programa “Nametbl” encuentran más faltas en el mismo que en “Cmdline” y “Ntree”. Con respecto, a la interacción del programa elegido con el programa aplicado, cuando los sujetos responden “Cmdline” encuentran menos faltas en él, que en “Nametbl” pero no que en “Ntree”; cuando responden “Ntree” encuentran menos faltas en el mismo y en “Cmdline” que en “Nametbl”. Además no influyen ni las preferencias que se tienen sobre las programas.

Debido a que en dos iteraciones en el mismo programa (“Cmdline”) se detectan menos faltas, se podría decir, que realmente en ciertos programas es más fácil encontrar las faltas que en otros. Finalmente, las preferencias están acorde a lo que responden, además, parece que es más probable acertar con las respuestas que disgustan que con las que gustan.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 1	,586	N/A	,263	N/A	,184	N/A
Iteración 2	,002	(na=nt)>cm	,422	N/A	,654	N/A
Iteración 3	,000	(cm=nt)>na	,294	N/A	,018	cm<na
						cm=nt
						na=nt
						na cm=na=nt
						nt (cm=nt)<na

Tabla 6.164. Resumen

6.2.3. ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa que los sujetos consideran que han entendido mejor. Esta pregunta aparece en la segunda y tercera iteración; además únicamente se preguntó a los sujetos después del estudio empírico.

6.2.3.1. Iteración 2

En la Tabla 6.165, se puede observar que para los programas "Nametbl" y "Ntree" coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de "Cmdline" el programa en el que han encontrado más faltas es "Nametbl". A continuación se va a evaluar si esta coincidencia es significativa.

PP3	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	40,7400	23,73243	9
	Nametbl	62,9644	13,88983	9
	Ntree	49,9989	18,63539	9
Nametbl	Cmdline	41,6663	20,18452	16
	Nametbl	54,1675	12,91253	16
	Ntree	53,1256	22,12644	16
Ntree	Cmdline	35,8977	20,23739	13
	Nametbl	46,1531	13,87028	13
	Ntree	52,5638	20,23723	13

Tabla 6.165. Estadísticos Descriptivos

En la Tabla 6.166, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$).

Box's M ^a	9,386
F	,675
df1	12
df2	3487,950
Sig.	,777

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP3

Within Subjects Design: AppliedProg

Tabla 6.166. Test de igualdad de matrices de covarianzas

En la Tabla 6.167, se observa que también es posible asumir esfericidad ($\text{sig}>0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,939	2,149	2	,341	,942	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP3

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.167. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.168, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}\leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig}>0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	4628,246	2	2314,123	7,017	,002
AppliedProg * PP3	Sphericity Assumed	1053,794	4	263,448	,799	,530
Error(AppliedProg)	Sphericity Assumed	23084,903	70	329,784		

Tabla 6.168. Test univariado de efectos intra-sujetos

En la Tabla 6.169, se observa que el programa que han entendido mejor no es significativo ($\text{sig}>0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	254405,010	1	254405,010	643,539	,000
PP3	777,840	2	388,920	,984	,384
Error	13836,263	35	395,322		

Tabla 6.169. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.170, se puede observar que la media más alta tienen los sujetos que aplicaron el programa "Nametbl" y la más baja los que aplicaron el programa "Cmdline". Esto se puede comprobar gráficamente en la Figura 6.22.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	39,435	3,515	32,299	46,570
Nametbl	54,428	2,248	49,865	58,992
Ntree	51,896	3,458	44,876	58,917

Tabla 6.170. Medias estimadas

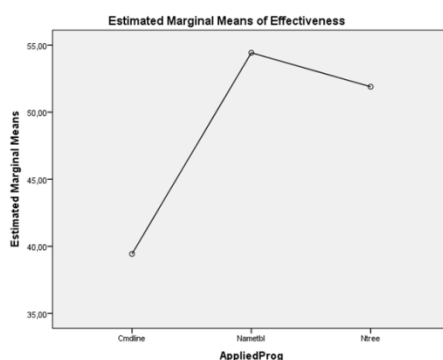


Figura 6.22. Programa Aplicado

Finalmente, en la Tabla 6.171, se puede comprobar que el programa “Nametbl” es significativamente mejor que “Cmdline”. Por otro lado, el programa “Ntree” también es significativamente mejor que la “Cmdline”. Finalmente no existen diferencias significativas entre el programa “Nametbl” y “Ntree”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-14,994*	3,991	,002	-25,028	-4,959
	Ntree	-12,461*	4,785	,040	-24,495	-,428
Nametbl	Cmdline	14,994*	3,991	,002	4,959	25,028
	Ntree	2,532	4,032	1,000	-7,606	12,670
Ntree	Cmdline	12,461*	4,785	,040	,428	24,495
	Nametbl	-2,532	4,032	1,000	-12,670	7,606

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.171. Comparación por pares

6.2.3.2. Iteración 3

En la Tabla 6.172, se puede observar que para el programa” Nametbl” coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de “Cmdline” el programa en el que han encontrado más faltas “Ntree” mientras que en el caso de “Ntree” el programa en el que han encontrado más faltas “Nametbl”. A continuación se va evaluar si esta coincidencia es significativa.

PP3	AppliedProg	Mean	Std. Desviation	N
Cmdline	Cmdline	44,1171	13,09973	17
	Nametbl	41,1765	21,34151	17
	Ntree	48,0388	21,95521	17
Nametbl	Cmdline	27,2718	13,48231	11
	Nametbl	54,5464	19,84852	11
	Ntree	39,3927	13,48501	11
Ntree	Cmdline	41,1765	21,34151	17
	Nametbl	50,9800	14,99075	17
	Ntree	39,2159	26,96545	17

Tabla 6.172. Estadísticos Descriptivos

En la Tabla 6.173, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig} > 0.05$).

Box's M ^a	16,842
F	1,248
df1	12
df2	5611,006
Sig.	,243

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP3

Within Subjects Design: AppliedProg

Tabla 6.173. Test de igualdad de matrices de covarianzas

En la Tabla 6.174, se puede observar que también es posible asumir esfericidad ($\text{sig} > 0.05$).

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,962	1,601	2	,449	,963	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP3

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.174. Test de esfericidad de Mauchly

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.175, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig} \leq 0.05$). Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	6061,240	2	3030,620	8,235	,001
AppliedProg * PP3	Sphericity Assumed	1328,937	4	332,234	,903	,466
Error(AppliedProg)	Sphericity Assumed	30914,734	84	368,033		

Tabla 6.175. Test univariado de efectos intra-sujetos

En la Tabla 6.176, se observa que las opiniones del programa que han entendido mejor no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	256521,780	1	256521,780	582,626	,000
PP3	1734,547	2	867,273	1,970	,152
Error	18491,980	42	440,285		

Tabla 6.176. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas

- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.177, se puede observar que la media más alta tienen los sujetos que aplicaron el programa “Nametbl” y la más baja los que aplicaron el programa “Cmdline”. Esto se puede comprobar gráficamente en la Figura 6.23.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	37,522	2,557	32,361	42,682
Nametbl	53,803	3,007	47,734	59,872
Ntree	42,216	3,417	35,320	49,112

Tabla 6.177. Medias estimadas

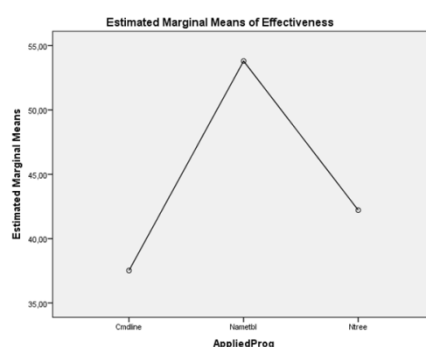


Figura 6.23. Programa Aplicado

Finalmente, en la Tabla 6.178, se puede comprobar que el programa “Nametbl” es significativamente mejor que “Cmdline” y “Ntree”. Finalmente no existen diferencias significativas entre el programa “Nametbl” y “Ntree”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-16,281*	3,996	,001	-26,246	-6,316
	Ntree	-4,694	3,856	,691	-14,309	4,921
Nametbl	Cmdline	16,281*	3,996	,001	6,316	26,246
	Ntree	11,587*	4,509	,041	,342	22,832
Ntree	Cmdline	4,694	3,856	,691	-4,921	14,309
	Nametbl	-11,587*	4,509	,041	-22,832	-,342

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.178. Comparación por pares

6.2.3.3. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente se ponen los datos que han salido significativos. En la Tabla 6.179, se observa que en la iteración 2, cuando los sujetos aplican el programa “Cmdline” encuentran menos faltas en el mismo que en “Nametbl” y “Ntree”. Además no

influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Finalmente, en la iteración 3 cuando los sujetos aplican el programa “Nametbl” encuentran más faltas en el mismo que en “Cmdline” y “Ntree”. Además no influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Debido a que en dos iteraciones en el mismo programa (“Nametbl”) se detectan más faltas, se podría decir, que realmente en ciertos programas es más fácil encontrar las faltas que en otros.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 2	,002	(na=nt)>cm	,384	N/A	,530	N/A
Iteración 3	,001	(cm=nt)<na	,152	N/A	,466	N/A

Tabla 6.179. Resumen

6.2.4. ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre la realidad (porcentaje de faltas encontradas al aplicar cada una de las técnicas) y el programa que los sujetos consideran que tiene las faltas más fáciles. Esta pregunta aparece únicamente en la tercera iteración; además únicamente se preguntó a los sujetos después del estudio empírico.

6.2.4.1. Iteración 3

En la Tabla 6.180, se puede observar que para el programa “Nametbl” coincide el programa elegido con el programa en el que han encontrado más faltas, en el caso de “Cmdline” el programa en el que han encontrado más faltas es “Nametbl” mientras que en el caso de “Ntree” el programa en el que han encontrado más faltas es “Cmdline”. A continuación se va a evaluar si esta coincidencia es significativa.

PP1	AppliedProg	Mean	Std. Deviation	N
Cmdline	Cmdline	39,1665	17,33339	20
	Nametbl	55,8335	22,47450	20
	Ntree	44,1660	23,11636	20
Nametbl	Cmdline	31,3718	20,31083	17
	Nametbl	56,8629	15,65663	17
	Ntree	41,1765	22,91094	17
Ntree	Cmdline	48,1478	13,03154	9
	Nametbl	44,4444	16,66792	9
	Ntree	42,5922	20,60209	9

Tabla 6.180. Estadísticos Descriptivos

En la Tabla 6.181, se observa que se puede asumir que las matrices de covarianzas son iguales ($\text{sig}>0.05$). En la Tabla 6.182, se observa que también es posible asumir esfericidad ($\text{sig}>0.05$).

El siguiente paso es analizar en los test univariados de los efectos intra-sujetos. Con respecto al programa aplicado, en la Tabla 6.183, se puede observar que para el programa aplicado el porcentaje de faltas encontradas por los sujetos al aplicar cada técnica no es el mismo ($\text{sig}\leq 0.05$).

Por otro lado, con respecto a la interacción del programa que los sujetos aplicaron y el programa que consideran más fácil, se puede observar que la influencia no es significativa ($\text{sig} > 0.05$).

Box's M ^a	9,971
F	,730
df1	12
df2	3230,629
Sig.	,723

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

Tabla 6.181. Test de igualdad de matrices de covarianzas

Within Subjects Effect ^a	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon ^b		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
AppliedProg	,976	1,003	2	,605	,977	1,000	,500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + PP1

Within Subjects Design: AppliedProg

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tabla 6.182. Test de esfericidad de Mauchly

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
AppliedProg	Sphericity Assumed	3663,686	2	1831,843	5,032	,009
AppliedProg * PP1	Sphericity Assumed	2519,479	4	629,870	1,730	,151
Error(AppliedProg)	Sphericity Assumed	31310,000	86	364,070		

Tabla 6.183. Test univariado de efectos intra-sujetos

En la Tabla 6.184, se observa que las opiniones del programa con las faltas más fáciles no son significativas ($\text{sig} > 0.05$).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	247078,626	1	247078,626	529,994	,000
PP1	292,075	2	146,037	,313	,733
Error	20046,230	43	466,191		

Tabla 6.184. Test de efectos entre-sujetos

De los resultados obtenidos se puede concluir que:

- Hay programas en los que se detectan más faltas
- Las preferencias de los sujetos sobre un programa no hace que se detecten más faltas
- Las preferencias no sirven para predecir. Por lo tanto, no garantizan que en un programa se vayan a encontrar más faltas que en otro.

Con respecto al programa aplicado, en la Tabla 6.185, se puede observar que la media más alta tienen los sujetos que aplicaron el programa "Nametbl" y la más baja los que aplicaron el programa "Cmdline". Esto se puede comprobar gráficamente en la Figura 6.24.

AppliedProg	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Cmdline	39,562	2,787	33,942	45,183
Nametbl	52,380	2,991	46,348	58,412
Ntree	42,645	3,532	35,523	49,767

Tabla 6.185. Medias estimadas

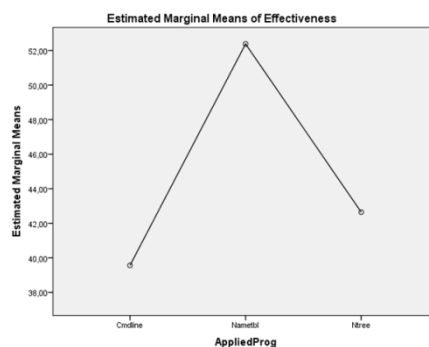


Figura 6.24. Programa Aplicado

Finalmente, en la Tabla 6.186, se observa que “Nametbl” es significativamente mejor que “Cmdline”. Y no hay diferencias entre “Cmdline” y “Ntree”, ni tampoco entre “Nametbl” y “Ntree”.

(I) AppliedTec	(J) AppliedTec	Mean Difference (I-J)	Std. Error	Sig. ^b	95% Confidence Interval for Difference ^b	
					Lower Bound	Upper Bound
Cmdline	Nametbl	-12,818*	4,048	,009	-22,903	-2,734
	Ntree	-3,083	4,058	1,000	-13,193	7,027
Nametbl	Cmdline	12,818*	4,048	,009	2,734	22,903
	Ntree	9,735	4,531	,112	-1,552	21,023
Ntree	Cmdline	3,083	4,058	1,000	-7,027	13,193
	Nametbl	-9,735	4,531	,112	-21,023	1,552

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Tabla 6.186. Comparación por pares

6.2.4.2. Discusión

Para facilitar la lectura, en las tablas resumen, únicamente aparecen los datos significativos. En la Tabla 6.187, se observa que, cuando los sujetos aplican el programa “Nametbl” encuentran más faltas en el mismo que en “Cmdline” pero no que en “Ntree”. Además no influyen ni las preferencias que se tienen sobre las programas, ni la interacción del programa elegido con el programa aplicado.

Debido a que en “Nametbl” se detectan más faltas, se podría decir, que realmente en ciertos programas es más fácil encontrar las faltas que en otros.

Iteración	Realidad	Pares	Elección	Pares	Realidad * Elección	Pares
Iteración 3	,009	na>cm	,733	N/A	,151	N/A
		cm=nt				
		na=nt				

Tabla 6.187. Resumen

CAPÍTULO 7. COMPARANDO LA INTUICIÓN CON LAS PREFERENCIAS

En este capítulo se va a analizar la relación existente entre las preguntas de intuición con las de preferencias. Para ello se seguirá el método explicado en el capítulo 2.

En esta sección se quiere dar respuesta a la pregunta: ¿qué subjetividad subyace en la intuición de los testers sobre la efectividad de las técnicas?

Para ello, se estudia si ha habido concordancias para cada sujeto entre lo que intuían y sus preferencias. Esto se hace comparando cada variable de la fila de la izquierda con las variables que aparecen en la derecha:

Resultados sobre Intuición	Resultados sobre Preferencias
Técnica más efectiva	Técnica preferida
Técnica mejor aplicada	Técnica más fácil de entender
	Técnica más fácil de aplicar
Programa más efectivo	Programa más fácil
	Programa más entendible
Programa menos efectivo	Programa con faltas más fáciles
	Programa con faltas más difíciles

Tabla 7.1. Preguntas de: Intuiciones vs Preferencias

7.1. Preguntas relacionadas con las Técnicas

7.1.1. ¿Qué técnica detecta más faltas? / ¿Qué técnica te gusta más?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que detecte más faltas y la técnica que más les gusta. Esta pregunta aparece en las tres iteraciones. Dado que ambas preguntas se realizaron antes y después de la realización del estudio empírico, se va a analizar las combinaciones: Antes-Antes, Después-Después.

7.1.1.1. Estudio del Acuerdo Antes-Antes

En la Figura 7.1, para la iteración 1, se observa que de cuatro personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, una también la considera su técnica preferida, de los tres restantes: para dos su preferida es “Estructural” y el otro es “Funcional”. De ocho personas que dicen que la técnica más efectiva es “Estructural” cinco también la consideran su técnica preferida, de los tres restantes: uno prefiere “Revisión” y los otros dos “Funcional”. Finalmente, de diez personas que dicen que la técnica más efectiva es “Funcional” cinco también la consideran su preferida y los otros cinco restantes tienen como técnica preferida “Estructural”.

Para la iteración 2, de dieciocho personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, dos también la consideran su técnica preferida, de los dieciséis restantes: para tres su preferida es “Estructural” y para trece la técnica preferida es “Funcional”. De doce personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran su técnica preferida, de los diez restantes: cuatro prefieren “Revisión” y los otros seis “Funcional”. Finalmente, de doce personas que dicen que la técnica más efectiva es “Funcional” diez también la consideran su preferida y los otros dos restantes tienen como técnica preferida “Revisión”.

Para la iteración 3, de dieciocho personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, siete también la consideran su técnica preferida, de los once restantes: para tres su preferida es “Estructural” y para ocho la técnica preferida es “Funcional”. De dieciséis personas que dicen que la técnica más efectiva es “Estructural” siete también la consideran su técnica preferida, de los nueve restantes: dos prefieren “Revisión” y los otros siete “Funcional”. Finalmente, de siete personas que dicen que la técnica más efectiva es “Funcional” cuatro también la consideran su preferida, de los tres restantes: para uno su favorita es “Revisión” y para dos la técnica preferida es “Estructural”.

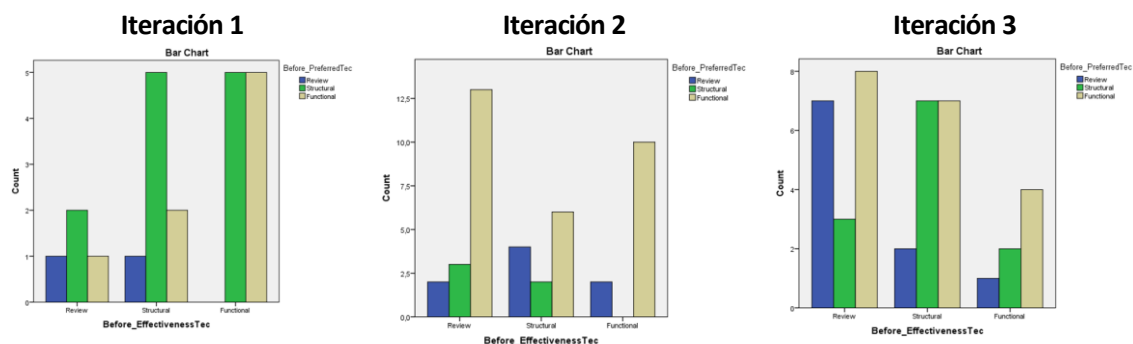


Figura 7.1. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.2, en la Tabla 7.3 y en la Tabla 7.4 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PT1(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	1	2	1	4
		% of Total	4,5%	9,1%	4,5%	18,2%
	Structural	Count	1	5	2	8
		% of Total	4,5%	22,7%	9,1%	36,4%
	Functional	Count	0	5	5	10
		% of Total	0,0%	22,7%	22,7%	45,5%
Total		Count	2	12	8	22
		% of Total	9,1%	54,5%	36,4%	100,0%

Tabla 7.2. Tabla de Contingencia de la iteración 1

			PT1(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	2	3	13	18
		% of Total	4,8%	7,1%	31,0%	42,9%
	Structural	Count	4	2	6	12
		% of Total	9,5%	4,8%	14,3%	28,6%
	Functional	Count	2	0	10	12
		% of Total	4,8%	0,0%	23,8%	28,6%
Total		Count	8	5	29	42
		% of Total	19,0%	11,9%	69,0%	100,0%

Tabla 7.3. Tabla de Contingencia de la iteración 2

			PT1(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	7	3	8	18
		% of Total	17,1%	7,3%	19,5%	43,9%
	Structural	Count	2	7	7	16
		% of Total	4,9%	17,1%	17,1%	39,0%
	Functional	Count	1	2	4	7
		% of Total	2,4%	4,9%	9,8%	17,1%
Total		Count	10	12	19	41
		% of Total	24,4%	29,3%	46,3%	100,0%

Tabla 7.4. Tabla de Contingencia de la iteración 3

7.1.1.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 49.9%, es decir, casi la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 33.4%, es decir, muy pocos sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 44%, es decir, menos de la mitad de los sujetos están de acuerdo.

En la Tabla 7.5, para las iteraciones 1 y 2, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo. Finalmente, para la iteración 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el resultado de Kappa muestra que el acuerdo es insignificante ($> 0,00 - 0,20$), se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,193	,165	1,243	,214
Iteración 2	,030	,083	,337	,736
Iteración 3	,198	,102	2,040	,041

Tabla 7.5. Índice de Concordancia Kappa

Debido a que para las iteraciones 1 y 2, no hay acuerdo, y que para la iteración 3, el acuerdo es insignificante, se puede concluir que los sujetos no identifican la técnica que detecta más faltas con la técnica que más les gusta.

7.1.1.2. Estudio del Acuerdo Después-Después

En la Figura 7.2, para la iteración 1, se observa que de cuatro personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, ninguna la considera su técnica preferida, por lo tanto para dos personas su favorita es “Estructural” y para las otras dos la técnica preferida es “Funcional”. De diez personas que dicen que la técnica más efectiva es “Estructural” ninguna la consideran su técnica preferida, por lo tanto, uno prefiere “Revisión” y los otros nueve “Funcional”. Finalmente, de ocho personas que dicen que la técnica más efectiva es “Funcional” cinco también la consideran su preferida y tres personas tienen como técnica preferida “Estructural”.

Para la iteración 2, de doce personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, tres también la consideran su técnica preferida, de los nueve restantes: para dos su preferida es “Estructural” y para siete la técnica preferida es “Funcional”. De once personas que dicen que la técnica más efectiva es “Estructural” ninguna la considera su técnica preferida, todos los sujetos tienen como favorita la técnica “Funcional”. Finalmente, de once personas que dicen que la técnica más efectiva es “Funcional” siete también la consideran su preferida de los cuatro restantes: para tres su favorita es “Revisión” y para uno la técnica preferida es “Estructural”.

Para la iteración 3, de dieciocho personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, siete también la consideran su técnica preferida, de los once restantes: para uno su preferida es “Estructural” y para diez la técnica preferida es “Funcional”. De diecisiete personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran su técnica preferida, de los quince restantes: cuatro prefieren “Revisión” y los once siete “Funcional”. Finalmente, de siete personas que dicen que la técnica más efectiva es “Funcional” cuatro también la consideran su preferida, de los tres restantes: para dos su favorita es “Revisión” y para uno la técnica preferida es “Estructural”.

Estos resultados se pueden contrastar en la Tabla 7.6, en la Tabla 7.7 y en la Tabla 7.8 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.



Figura 7.2. Acuerdo entre las respuestas

			PT1(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	0	2	2	4
		% of Total	0,0%	9,1%	9,1%	18,2%
	Structural	Count	1	0	9	10
		% of Total	4,5%	0,0%	40,9%	45,5%
	Functional	Count	0	3	5	8
		% of Total	0,0%	13,6%	22,7%	36,4%
Total		Count	1	5	16	22
		% of Total	4,5%	22,7%	72,7%	100,0%

Tabla 7.6. Tabla de Contingencia de la iteración 1

			PT1(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	5	1	8	14
		% of Total	13,5%	2,7%	21,6%	37,8%
	Structural	Count	3	2	2	7
		% of Total	8,1%	5,4%	5,4%	18,9%
	Functional	Count	1	0	15	16
		% of Total	2,7%	0,0%	40,5%	43,2%
Total		Count	9	3	25	37
		% of Total	24,3%	8,1%	67,6%	100,0%

Tabla 7.7. Tabla de Contingencia de la iteración 2

			PT1(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	12	0	8	20
		% of Total	26,7%	0,0%	17,8%	44,4%
	Structural	Count	0	2	4	6
		% of Total	0,0%	4,4%	8,9%	13,3%
	Functional	Count	2	2	15	19
		% of Total	4,4%	4,4%	33,3%	42,2%
Total		Count	14	4	27	45
		% of Total	31,1%	8,9%	60,0%	100,0%

Tabla 7.8. Tabla de Contingencia de la iteración 3

7.1.1.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 22.7%, es decir, muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 59.4%, es decir, más de la mitad de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 64.4%, es decir, más de la mitad de los sujetos están de acuerdo.

En la Tabla 7.9, para la iteración 1, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo. Para las iteraciones 2 y 3, el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el resultado de Kappa muestra que el acuerdo es discreto (0,21 - 0,40), se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	-,238	,115	-1,763	,078
Iteración 2	,325	,116	2,843	,004
Iteración 3	,404	,113	3,613	,000

Tabla 7.9. Índice de Concordancia Kappa

Debido a que para la iteración 1, no hay acuerdo, y que para las iteraciones 2 y 3, el acuerdo es discreto, se puede concluir que parece que los sujetos, al cambiar de opinión, identifican la técnica que detecta más faltas con la técnica que más les gusta. La relación entre las preguntas es baja.

7.1.2. ¿Qué técnica detecta más faltas? / ¿Qué técnica te ha parecido más fácil de entender?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que detecte más faltas y la técnica que creen que han entendido mejor o que es más fácil de entender. Esta pregunta aparece en las tres iteraciones. Dado que la segunda pregunta solo se realizó después de la realización del estudio empírico, se van a analizar únicamente las respuestas: Después-Después.

7.1.2.1. Estudio del Acuerdo Después-Después

En la Figura 7.3, para la iteración 1, se observa que de cuatro personas que dicen, después de aplicar las técnicas, que la técnica más efectiva es "Revisión", ninguna la considera la técnica más fácil de entender, por lo tanto para dos personas la más fácil de entender es "Estructural" y para las otras dos la más fácil de entender es "Funcional". De ocho personas que dicen que la técnica más efectiva es "Estructural" dos también la consideran la técnica más fácil de entender, de los seis restantes: para uno la más fácil de entender es "Revisión" y los otros cinco "Funcional". Finalmente se observa que de diez personas que dicen que la técnica más efectiva es "Funcional" nueve también la consideran la técnica más fácil de entender y solo una persona cree que la más fácil de entender es "Estructural".

Para la iteración 2, de doce personas que dicen, después de aplicar las técnicas, que la técnica más efectiva es “Revisión”, tres también la consideran la técnica más fácil de entender, de los nueve restantes: para dos su favorita es “Estructural” y para siete la técnica preferida es “Funcional”. De once personas que dicen que la técnica más efectiva es “Estructural” también la consideraban la técnica más fácil de entender, todos los sujetos tienen como favorita la técnica “Funcional”. Finalmente se observa que de once personas que dicen que la técnica más efectiva es “Funcional” siete también la consideraban la técnica más fácil de entender de los cuatro restantes: para tres su favorita es “Revisión” y para uno la técnica preferida es “Estructural”.

Para la iteración 3, de dieciocho personas que dicen, después de aplicar las técnicas, que la técnica más efectiva es “Revisión”, siete también la consideraban la técnica más fácil de entender, de los once restantes: para uno su favorita es “Estructural” y para diez la técnica preferida es “Funcional”. De diecisiete personas que dicen que la técnica más efectiva es “Estructural” dos también la consideraban la técnica más fácil de entender, de los quince restantes: cuatro prefieren “Revisión” y los once siete “Funcional”. Finalmente se observa que de siete personas que dicen que la técnica más efectiva es “Funcional” cuatro también la consideraban la técnica más fácil de entender, de los tres restantes: para dos su favorita es “Revisión” y para uno la técnica preferida es “Estructural”.

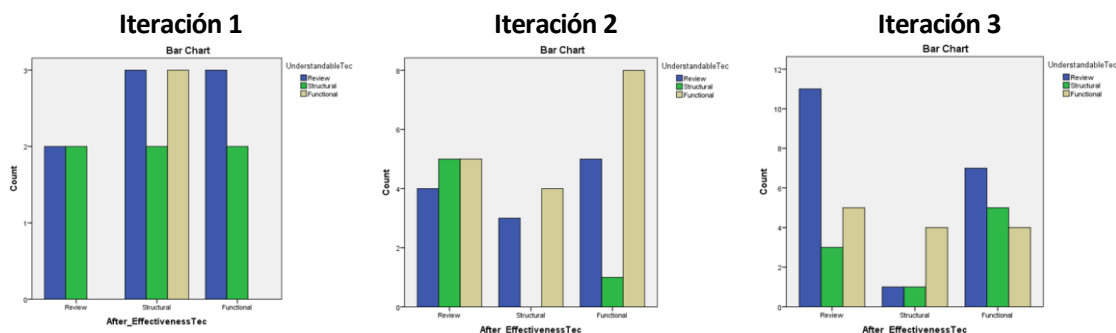


Figura 7.3. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.10, en la Tabla 7.11 y en la Tabla 7.12 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PT2			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	2	2	0	4
		% of Total	11,8%	11,8%	0,0%	23,5%
	Structural	Count	3	2	3	8
		% of Total	17,6%	11,8%	17,6%	47,1%
	Functional	Count	3	2	0	5
		% of Total	17,6%	11,8%	0,0%	29,4%
Total		Count	8	6	3	17
		% of Total	47,1%	35,3%	17,6%	100,0%

Tabla 7.10. Tabla de Contingencia de la iteración 1

			PT2			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	4	5	5	14
		% of Total	28,6%	35,7%	35,7%	40,0%
	Structural	Count	3	0	4	7
		% of Total	42,9%	0,0%	57,1%	20,0%
	Functional	Count	5	1	8	14
		% of Total	35,7%	7,2%	57,1%	40,0%
Total		Count	12	6	17	35
		% of Total	34,3%	17,1%	48,6%	100,0%

Tabla 7.11. Tabla de Contingencia de la iteración 2

			PT2			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	11	3	5	19
		% of Total	26,8%	7,3%	12,2%	46,3%
	Structural	Count	1	1	4	6
		% of Total	2,4%	2,4%	9,8%	14,6%
	Functional	Count	7	5	4	16
		% of Total	17,1%	12,2%	9,8%	39,0%
Total		Count	19	9	13	42
		% of Total	46,3%	22,0%	31,7%	100,0%

Tabla 7.12. Tabla de Contingencia de la iteración 3

7.1.2.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 23.6%, es decir, muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 85.7%, es decir, la mayoría de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 39%, es decir, muy pocos sujetos están de acuerdo.

En la Tabla 7.13, para las tres iteraciones, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	-,139	,142	-,867	,386
Iteración 2	-,036	,110	-,293	,769
Iteración 3	,031	,108	,275	,784

Tabla 7.13. Índice de Concordancia Kappa

Debido a que en ninguna de las tres iteraciones hay acuerdo, se puede concluir que los sujetos no identifican la técnica que detecta más faltas con la técnica que les ha parecido más fácil de entender.

7.1.3. ¿Qué técnica detecta más faltas? / ¿Qué técnica te ha parecido más fácil de aplicar?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que detecte más faltas y la técnica que creen es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones. Dado que ambas preguntas se realizaron antes y después de la realización del estudio empírico, se va a analizar las combinaciones: Antes-Antes, Después-Después.

7.1.3.1. Estudio del Acuerdo Antes-Antes

En la Figura 7.4, para la iteración 1, se observa que de cuatro personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, dos también la consideraban la técnica más fácil de aplicar, de los dos restantes: para uno la técnica más fácil de aplicar “Estructural” y para otro es “Funcional”. De nueve personas que dicen que la técnica más efectiva es “Estructural” tres también la consideran la técnica más fácil de aplicar, de los seis restantes: dos consideran la técnica más fácil de aplicar “Revisión” y los otros cuatro “Funcional”. Finalmente se observa que de diez personas que dicen que la técnica más efectiva es “Funcional” siete también la consideran la más fácil de aplicar, de los tres restantes: dos consideraban que la más fácil de aplicar “Revisión” y otro “Estructural”.

Para la iteración 2, de dieciocho personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, dos también la consideran la técnica más fácil de aplicar, de los dieciséis restantes: para tres la técnica más fácil de aplicar “Estructural” y para trece es “Funcional”. De doce personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran la técnica más fácil de aplicar, de los diez restantes: cuatro consideraban la técnica más fácil de aplicar “Revisión” y los otros seis “Funcional”. Finalmente se observa que de doce personas que dicen que la técnica más efectiva es “Funcional” diez también la consideran la técnica más fácil de aplicar y los dos restantes consideraban que “Revisión” es la técnica más fácil de aplicar.

Para la iteración 3, de dieciocho personas que dicen, antes de aplicar las técnicas, que la técnica más efectiva es “Revisión”, dos también la consideran la técnica más fácil de aplicar, de los dieciséis restantes: para seis la técnica más fácil de aplicar es “Estructural” y para diez es “Funcional”. De dieciséis personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran la técnica más fácil de aplicar, de los catorce restantes: seis consideran la técnica más fácil de aplicar “Revisión” y los otros once “Funcional”. Finalmente se observa que de siete personas que dicen que la técnica más efectiva es “Funcional” tres también la consideran la técnica más fácil de aplicar, de los cuatro restantes: para dos la técnica más fácil de aplicar es “Revisión” y para los otros dos “Estructural”.

Estos resultados se pueden contrastar en la Tabla 7.14, en la Tabla 7.15 y en la Tabla 7.16 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a qué porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

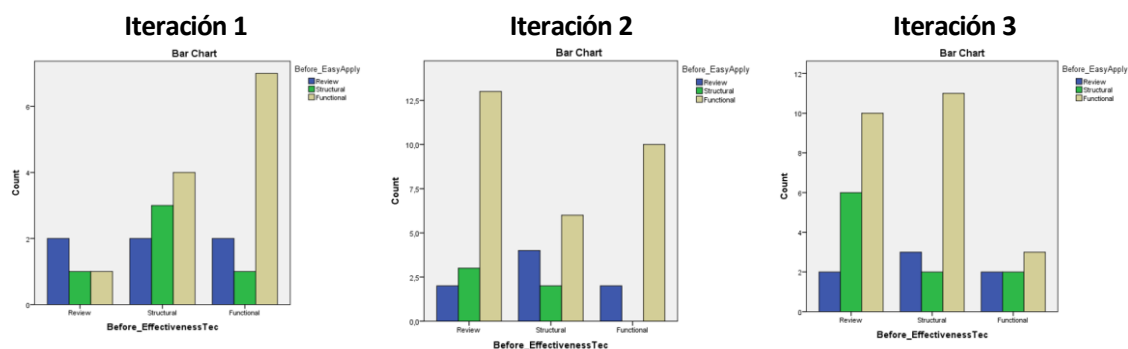


Figura 7.4. Acuerdo entre las respuestas

			PT3(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	2	1	1	4
	Review	% of Total	8,7%	4,3%	4,3%	17,4%
	Structural	Count	2	3	4	9
	Structural	% of Total	8,7%	13,0%	17,4%	39,1%
	Functional	Count	2	1	7	10
	Functional	% of Total	8,7%	4,3%	30,4%	43,5%
Total		Count	6	5	12	23
		% of Total	26,1%	21,7%	52,2%	100,0%

Tabla 7.14. Tabla de Contingencia de la iteración 1

			PT3(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	2	3	13	18
	Review	% of Total	4,8%	7,1%	31,0%	42,9%
	Structural	Count	4	2	6	12
	Structural	% of Total	9,5%	4,8%	14,3%	28,6%
	Functional	Count	2	0	10	12
	Functional	% of Total	4,8%	0,0%	23,8%	28,6%
Total		Count	8	5	29	42
		% of Total	19,0%	11,9%	69,0%	100,0%

Tabla 7.15. Tabla de Contingencia de la iteración 2

			PT3(Before)			Total
			Review	Structural	Functional	
IT1(Before)	Review	Count	2	6	10	18
	Review	% of Total	4,9%	14,6%	24,4%	43,9%
	Structural	Count	3	2	11	16
	Structural	% of Total	7,3%	4,9%	26,8%	39,0%
	Functional	Count	2	2	3	7
	Functional	% of Total	4,9%	4,9%	7,3%	17,1%
Total		Count	7	10	24	41
		% of Total	17,1%	24,4%	58,5%	100,0%

Tabla 7.16. Tabla de Contingencia de la iteración 3

7.1.3.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 52.1%, es decir, poco más de la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 33.4%, es decir, muy pocos sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 17.1%, es decir, muy pocos sujetos están de acuerdo.

En la Tabla 8.17, para las tres iteraciones, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,256	,151	1,780	,075
Iteración 2	,030	,083	,337	,736
Iteración 3	-,136	,080	-1,583	,113

Tabla 7.17. Índice de Concordancia Kappa

Debido a que en ninguna de las tres iteraciones hay acuerdo, se puede concluir que los sujetos no identifican la técnica que detecta más faltas con la técnica más fácil de aplicar.

7.1.3.2. Estudio del Acuerdo Después-Después

En la Figura 7.5, para la iteración 1, se observa que de cuatro personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, ninguna la considera la técnica más fácil de aplicar, por lo tanto para dos personas la más fácil de aplicar es “Estructural” y para las otras dos es “Funcional”. De diez personas que dicen que la técnica más efectiva es “Estructural” ninguna la considera la técnica más fácil de aplicar, por lo tanto, para uno la más fácil de aplicar es “Revisión” y los otros nueve “Funcional”. Finalmente, de ocho personas que dicen que la más fácil de aplicar es “Funcional” para cinco también la más fácil de aplicar y los otros tres es “Estructural”.

Para la iteración 2, de doce personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, tres también la consideran la técnica más fácil de aplicar, de los nueve restantes: para dos la más fácil de aplicar es “Estructural” y para siete la más fácil de aplicar es “Funcional”. De once personas que dicen que la técnica más efectiva es “Estructural” ninguna la considera más fácil de aplicar, todos los sujetos consideran que la más fácil de aplicar es “Funcional”. Finalmente, de once personas que dicen que la técnica más efectiva es “Funcional” siete también la consideran la más fácil de aplicar de los cuatro restantes: para tres la más fácil de aplicar es “Revisión” y para uno la más fácil de aplicar es “Estructural”.

Para la iteración 3, de dieciocho personas que dicen, después de la aplicación de las técnicas, que la técnica más efectiva es “Revisión”, siete también la consideran su técnica preferida, de los once restantes: para uno su preferida es “Estructural” y para diez es “Funcional”. De diecisiete personas que dicen que la técnica más efectiva es “Estructural” dos también la consideran su técnica preferida, de los quince restantes: cuatro prefieren “Revisión” y los once siete “Funcional”. Finalmente, de siete personas que dicen que la técnica más efectiva es “Funcional” cuatro también la consideran la más fácil de aplicar, de los tres restantes: para dos la más fácil de aplicar es “Revisión” y para el otro es “Estructural”.

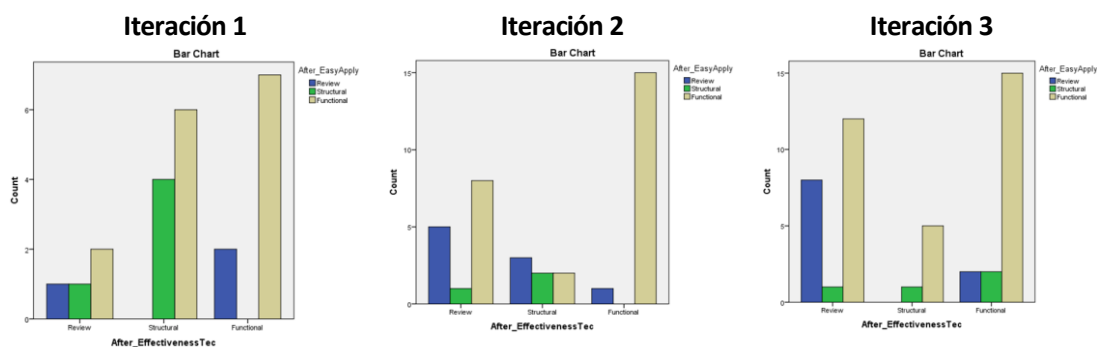


Figura 7.5. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.18, Tabla 7.19 y Tabla 7.20 para las iteraciones 1, 2 y 3, respectivamente. Donde se indica el porcentaje de sujetos de cada técnica.

			PT3(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	1	1	2	4
		% of Total	4,3%	4,3%	8,7%	17,4%
	Structural	Count	0	4	6	10
		% of Total	0,0%	17,4%	26,1%	43,5%
	Functional	Count	2	0	7	9
		% of Total	8,7%	0,0%	30,4%	39,1%
Total		Count	3	5	15	23
		% of Total	13,0%	21,7%	65,2%	100,0%

Tabla 7.18. Tabla de Contingencia de la iteración 1

			PT3(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	5	1	8	14
		% of Total	13,5%	2,7%	21,6%	37,8%
	Structural	Count	3	2	2	7
		% of Total	8,1%	5,4%	5,4%	18,9%
	Functional	Count	1	0	15	16
		% of Total	2,7%	0,0%	40,5%	43,2%
Total		Count	9	3	25	37
		% of Total	24,3%	8,1%	67,6%	100,0%

Tabla 7.19. Tabla de Contingencia de la iteración 2

			PT3(After)			Total
			Review	Structural	Functional	
IT1(After)	Review	Count	8	1	12	21
		% of Total	17,4%	2,2%	26,1%	45,7%
	Structural	Count	0	1	5	6
		% of Total	0,0%	2,2%	10,9%	13,0%
	Functional	Count	2	2	15	19
		% of Total	4,3%	4,3%	32,6%	41,3%
Total		Count	10	4	32	46
		% of Total	21,7%	8,7%	69,6%	100,0%

Tabla 7.20. Tabla de Contingencia de la iteración 3

7.1.3.2.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 52.1%, es decir, poco más de la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 59.4%, es decir, la mayoría de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 52.2%, es decir, poco más de la mitad de los sujetos están de acuerdo.

En la Tabla 7.21, para la iteración 1, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo. Para las iteraciones 2 y 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor de Kappa muestra que el acuerdo es discreto (0,21 - 0,40), se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,238	,146	1,706	,088
Iteración 2	,325	,116	2,843	,004
Iteración 3	,206	,105	2,047	,041

Tabla 7.21. Índice de Concordancia Kappa

Debido a que en la iteración 1 no hay acuerdo, y que en las iteraciones 2 y 3 hay un acuerdo discreto, se puede concluir que parece que los sujetos, al cambiar de opinión, identifican la técnica que detecta más faltas con la técnica que les ha parecido más fácil de aplicar. Sin embargo no hay mucha relación.

7.1.4. ¿Qué técnica has aplicado mejor? / ¿Qué técnica te gusta más?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que detecte más faltas y la técnica que ellos consideran que la aplicaron mejor. Esta pregunta aparece en las tres iteraciones. Dado que la primera pregunta únicamente se realiza después del estudio empírico, se va a analizar la combinación: Después-Después.

7.1.4.1. Estudio del Acuerdo Después-Después

En la Figura 7.6, para la iteración 1, se observa que de tres personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron fue "Revisión", ninguna la considera su técnica preferida, por lo tanto para las tres personas su preferida es "Funcional". De ocho personas que dicen que la técnica que mejor aplicaron fue "Estructural" tres también la consideran su técnica preferida y los cinco restantes prefieren "Funcional". Finalmente, de trece personas que dicen que la técnica que mejor aplicaron fue "Funcional" nueve también la consideran su preferida, de los cuatro restantes: para uno su preferida es "Revisión" y para tres la técnica preferida es "Estructural".

Para la iteración 2, de diez personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron fue "Revisión", siete también la consideran su técnica preferida, pero los tres restantes la técnica preferida es "Funcional". De seis personas que dicen que la técnica que mejor aplicaron fue "Estructural" tres la consideran su técnica preferida, de los tres restantes: para uno su

preferida es “Revisión” y para dos la técnica preferida es “Funcional”. Finalmente, de veintidós personas que dicen que la técnica que mejor aplicaron fue “Funcional” veintiuno también la consideran su preferida y solo una persona prefiere “Revisión”.

Para la iteración 3, de dieciocho personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron fue “Revisión”, once también la consideran su técnica preferida y para siete la técnica preferida es “Funcional”. De cinco personas que dicen que la técnica que mejor aplicaron fue “Estructural” cuatro también la consideran su técnica preferida y solo una persona prefiere “Funcional”. Finalmente, de veintidós personas que dicen que la técnica que mejor aplicaron fue “Funcional” para diecinueve también es su preferida y para los tres es “Revisión”.

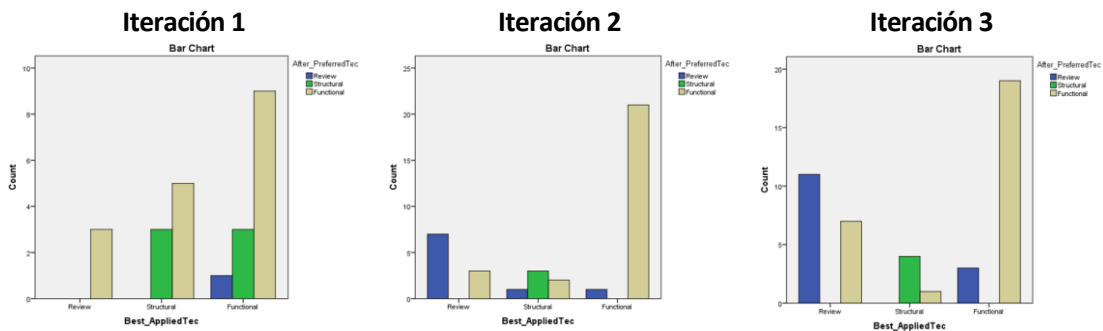


Figura 7.6. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.22, Tabla 7.23 y Tabla 7.24 para las iteraciones 1, 2 y 3, respectivamente. Donde se indica el porcentaje de sujetos en cada técnica.

			PT1(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	0	0	3	3
	Review	% of Total	0,0%	0,0%	12,5%	12,5%
	Structural	Count	0	3	5	8
	Structural	% of Total	0,0%	12,5%	20,8%	33,3%
	Functional	Count	1	3	9	13
	Functional	% of Total	4,2%	12,5%	37,5%	54,2%
Total		Count	1	6	17	24
		% of Total	4,2%	25,0%	70,8%	100,0%

Tabla 7.22. Tabla de Contingencia de la iteración 1

			PT1(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	7	0	3	10
	Review	% of Total	18,4%	0,0%	7,9%	26,3%
	Structural	Count	1	3	2	6
	Structural	% of Total	2,6%	7,9%	5,3%	15,8%
	Functional	Count	1	0	21	22
	Functional	% of Total	2,6%	0,0%	55,3%	57,9%
Total		Count	9	3	26	38
		% of Total	23,7%	7,9%	68,4%	100,0%

Tabla 7.23. Tabla de Contingencia de la iteración 2

			PT1(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	11	0	7	18
		% of Total	24,4%	0,0%	15,6%	40,0%
	Structural	Count	0	4	1	5
		% of Total	0,0%	8,9%	2,2%	11,1%
	Functional	Count	3	0	19	22
		% of Total	6,7%	0,0%	42,2%	48,9%
Total		Count	14	4	27	45
		% of Total	31,1%	8,9%	60,0%	100,0%

Tabla 7.24. Tabla de Contingencia de la iteración 3

7.1.4.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 50%, es decir, la mitad de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 81.6%, es decir, la mayoría de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 75.5%, es decir, la mayoría de los sujetos están de acuerdo

En la Tabla 7.25, para la iteración 1, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$) se puede concluir que no existe acuerdo. Para la iteración 2, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor de Kappa muestra que el acuerdo es sustancial (0,61 - 0,80), se puede concluir que existe acuerdo. Para las iteraciones 2 y 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor de Kappa muestra que el acuerdo es moderado ($> 0,41$ - 0,60), se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,053	,166	,334	,738
Iteración 2	,652	,115	5,345	,000
Iteración 3	,573	,114	4,910	,000

Tabla 7.25. Índice de Concordancia Kappa

Debido a que en la iteración 1 no hay acuerdo, y que en la iteración 2 hay un acuerdo sustancial y en la iteración 3 el acuerdo es moderado, se puede concluir que los sujetos, identifican la técnica que han aplicado mejor con la técnica que más les gusta. La relación existente es intermedia.

7.1.5. ¿Qué técnica has aplicado mejor? / ¿Qué técnica te ha parecido más fácil de entender?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que han aplicado mejor y la técnica que creen que han entendido mejor o que es más fácil de entender. Esta pregunta aparece en las tres iteraciones. Dado que ambas preguntas solo se realizaron después de la realización del estudio empírico, se va a analizar la única combinación: Después-Después.

7.1.5.1. Estudio del Acuerdo Después-Después

En la Figura 7.7, para la iteración 1, se observa que de dos personas que dicen, después de aplicar las técnicas, que la técnica mejor aplicada es “Revisión”, uno también la considera la técnica más fácil de entender, para la persona restante la más fácil de entender es “Funcional”. De cuatro personas que dicen que la técnica mejor aplicada es “Estructural” uno también la considera la más fácil de entender, de los tres restantes: dos consideran que la más fácil de entender es “Revisión” y uno “Funcional”. Finalmente, de diez personas que dicen que la técnica mejor aplicada es “Funcional” uno también la considera la técnica más fácil de entender de los nueve restantes: para cuatro la más fácil de entender es “Revisión” y para los otros cinco es “Estructural”.

Para la iteración 2, de diez personas que dicen, después de aplicar las técnicas, que la técnica mejor aplicada es “Revisión”, cinco también la consideran la técnica más fácil de entender, de los cinco restantes: para cuatro la más fácil de entender es “Estructural” y para uno la más fácil de entender es “Funcional”. De seis personas que dicen que la técnica mejor aplicada es “Estructural” ninguna la considera la técnica más fácil de entender, los sujetos se distribuyen de la siguiente manera: dos consideran que la más fácil de entender es “Revisión” y los otros cuatro “Funcional”. Finalmente, de veinte personas que dicen que la técnica mejor aplicada es “Funcional” trece también la consideraban la técnica más fácil de entender, de los siete restantes: para cinco la más fácil de entender es “Revisión” y para los otros dos es “Estructural”.

Para la iteración 3, de dieciséis personas que dicen, después de aplicar las técnicas, que la técnica mejor aplicada es “Revisión”, diez también la consideran la técnica más fácil de entender, de los seis restantes: para tres la más fácil de entender es “Estructural” y para tres la más fácil de entender es “Funcional”. De cuatro personas que dicen que la técnica mejor aplicada es “Estructural” una también la considera la técnica más fácil de entender, de los tres restantes: uno considera que la más fácil de entender es “Revisión” y los otros dos “Funcional”. Finalmente, de veintiuno personas que dicen que la técnica mejor aplicada es “Funcional” ocho también la consideraban la técnica más fácil de entender de los trece: para ocho la más fácil de entender es “Revisión” y para los otros cinco es “Estructural”.

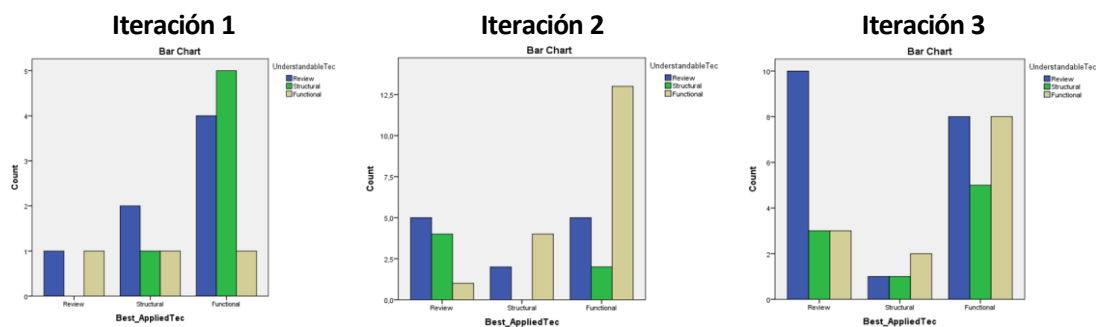


Figura 7.7. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.26, en la Tabla 7.27 y en la Tabla 7.28 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a qué porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

			PP2			Total
			Review	Structural	Funcional	
IT2	Review	Count	1	0	1	2
		% of Total	6,2%	0,0%	6,2%	12,5%
	Structural	Count	2	1	1	4
		% of Total	12,5%	6,2%	6,2%	25,0%
	Funcional	Count	4	5	1	10
		% of Total	25,0%	31,2%	6,2%	62,5%
Total		Count	7	6	3	16
		% of Total	43,8%	37,5%	18,8%	100,0%

Tabla 7.26. Tabla de Contingencia de la iteración 1

			PP2			Total
			Review	Structural	Functional	
IT2	Review	Count	5	4	1	10
		% of Total	13,9%	11,1%	2,8%	27,8%
	Structural	Count	2	0	4	6
		% of Total	5,6%	0,0%	11,1%	16,7%
	Functional	Count	5	2	13	20
		% of Total	13,9%	5,6%	36,1%	55,6%
Total		Count	12	6	18	36
		% of Total	33,3%	16,7%	50,0%	100,0%

Tabla 7.27. Tabla de Contingencia de la iteración 2

			PP2			Total
			Review	Structural	Functional	
IT2	Review	Count	10	3	3	16
		% of Total	24,4%	7,3%	7,3%	39,0%
	Structural	Count	1	1	2	4
		% of Total	2,4%	2,4%	4,9%	9,8%
	Functional	Count	8	5	8	21
		% of Total	19,5%	12,2%	19,5%	51,2%
Total		Count	19	9	13	41
		% of Total	46,3%	22,0%	31,7%	100,0%

Tabla 7.28. Tabla de Contingencia de la iteración 3

7.1.5.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 18.6%, es decir, muy pocos sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 50%, es decir, la mitad de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 46.3%, es decir, casi la mitad de los sujetos están de acuerdo

En la Tabla 7.29, para las tres iteraciones, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), por lo tanto, se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	-,106	,136	-,803	,422
Iteración 2	,169	,114	1,375	,169
Iteración 3	,155	,111	1,408	,159

Tabla 7.29. Índice de Concordancia Kappa

Debido a que en ninguna de las tres iteraciones hay acuerdo, se puede concluir que los sujetos no identifican la técnica que han aplicado mejor con la técnica que les parece más fácil de entender.

7.1.6. ¿Qué técnica has aplicado mejor? / ¿Qué técnica es más fácil de aplicar?

En esta sección se va a evaluar la relación existente entre la técnica que los sujetos consideran que han aplicado mejor y la técnica que creen es la más fácil de aplicar. Esta pregunta aparece en las tres iteraciones. Dado que la primera pregunta solo se realizó después de la realización del estudio empírico, se va a analizar la combinación: Después-Después.

7.1.6.1. Estudio del Acuerdo Después-Después

En la Figura 7.8, para la iteración 1, se observa que de tres personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron es “Revisión”, una también la considera la más fácil de aplicar, para los dos restantes la más fácil de aplicar es “Funcional”. De ocho personas que dicen que la técnica que mejor aplicaron es “Estructural” cuatro también la consideran técnica más fácil de aplicar, de los cuatro restantes: uno considera que la técnica más fácil de aplicar es “Revisión” y los otros tres “Funcional”. Finalmente, de quince personas que dicen que la técnica que mejor aplicaron es “Funcional” trece también la consideran la técnica más fácil de aplicar, de los dos restantes: para uno su favorita es “Revisión” y para otro es “Estructural”.

Para la iteración 2, de diez personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron es “Revisión”, siete también la consideran la técnica más fácil de aplicar y para los tres restantes la más fácil de aplicar es “Funcional”. De seis personas que dicen que la técnica que mejor aplicaron es “Estructural” tres la consideran la técnica más fácil de aplicar, los tres restantes: uno considera que la técnica más fácil de aplicar es “Revisión” y los otros dos “Funcional”. Finalmente, de veintidós personas que dicen que la técnica que mejor aplicaron era “Funcional” veintiuno también la consideran su preferida y para uno su favorita es “Revisión”.

Para la iteración 3, de diecinueve personas que dicen, después de aplicar las técnicas, que la técnica que mejor aplicaron es “Revisión”, ocho también la consideran la técnica más fácil de aplicar, de los once restantes: para uno la técnica más fácil de aplicar es “Estructural” y para diez es “Funcional”. De cinco personas que dicen que la técnica que mejor aplicaron es “Estructural” tres también la consideran la técnica más fácil de aplicar, para los dos restantes la técnica más fácil de aplicar es “Funcional”. Finalmente, de veintidós personas que dicen que la técnica que mejor aplicaron es “Funcional” veinte también la consideran la técnica más fácil de aplicar y para dos la técnica más fácil de aplicar es “Revisión”.

Estos resultados se pueden contrastar en la Tabla 7.30, en la Tabla 7.31 y en la Tabla 7.32 para las iteraciones 1, 2 y 3, respectivamente. Donde además se indica a qué porcentaje de sujetos corresponden las respuestas de cada una de las técnicas.

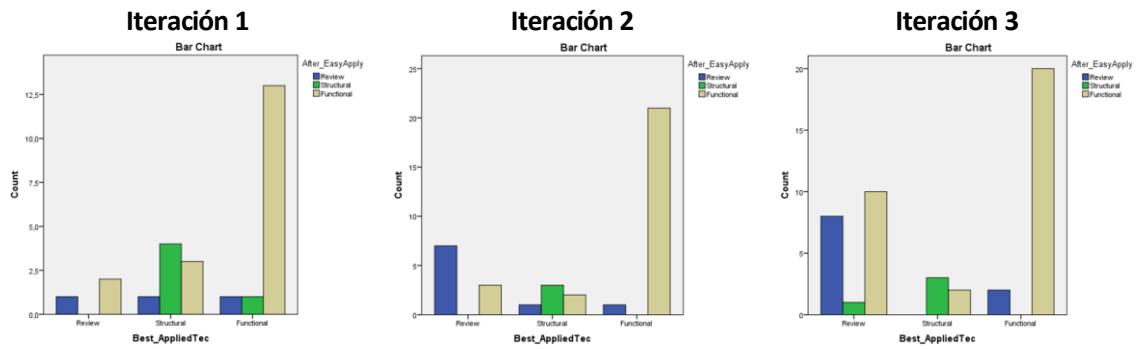


Figura 7.8. Acuerdo entre las respuestas

			PT3(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	1	0	2	3
		% of Total	3,8%	0,0%	7,7%	11,5%
	Structural	Count	1	4	3	8
		% of Total	3,8%	15,4%	11,5%	30,8%
	Functional	Count	1	1	13	15
		% of Total	3,8%	3,8%	50,0%	57,7%
Total	Count	3	5	18	26	
	% of Total	11,5%	19,2%	69,2%	100,0%	

Tabla 7.30. Tabla de Contingencia de la iteración 1

			PT3(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	7	0	3	10
		% of Total	18,4%	0,0%	7,9%	26,3%
	Structural	Count	1	3	2	6
		% of Total	2,6%	7,9%	5,3%	15,8%
	Functional	Count	1	0	21	22
		% of Total	2,6%	0,0%	55,3%	57,9%
Total		Count	9	3	26	38
		% of Total	23,7%	7,9%	68,4%	100,0%

Tabla 7.31. Tabla de Contingencia de la iteración 2

			PT3(After)			Total
			Review	Structural	Functional	
IT2	Review	Count	8	1	10	19
		% of Total	17,4%	2,2%	21,7%	41,3%
	Structural	Count	0	3	2	5
		% of Total	0,0%	6,5%	4,3%	10,9%
	Functional	Count	2	0	20	22
		% of Total	4,3%	0,0%	43,5%	47,8%
Total		Count	10	4	32	46
		% of Total	21,7%	8,7%	69,6%	100,0%

Tabla 7.32. Tabla de Contingencia de la iteración 3

7.1.6.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 1 es del 69.2%, es decir, la mayoría de los sujetos están de acuerdo. Para la iteración 2 el porcentaje es de 81.6%, es decir, la mayoría de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 67.4%, es decir, la mayoría de los sujetos están de acuerdo

En la Tabla 7.33, para las iteraciones 1 y 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$), y dado que el valor Kappa muestra que el acuerdo es moderado ($>0,41 - 0,60$), por lo tanto, se puede concluir que existe acuerdo.

Para la iteración 2, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$), y dado que el valor Kappa muestra que el acuerdo es sustancial ($0,61 - 0,80$), por lo tanto, se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 1	,417	,157	2,855	,004
Iteración 2	,652	,115	5,345	,000
Iteración 3	,426	,114	3,986	,000

Tabla 7.33. Índice de Concordancia Kappa

Debido a que en las iteraciones 1 y 3 hay un acuerdo moderado y en la iteración 3 el acuerdo es sustancial, se puede concluir que los sujetos identifican la técnica que han aplicado mejor con la técnica que les parece más fácil de aplicar. La relación existente es intermedia.

7.2. Preguntas relacionadas con los Programas

7.2.1. ¿Con qué programa has detectado más faltas?/ ¿Qué programa es más fácil?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron más faltas y el programa que consideran era el más fácil. Esta pregunta aparece en la segunda y tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.1.1. Estudio del Acuerdo Después-Después

En la Figura 7.9, para la iteración 2, de dieciséis personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron más faltas es "Cmdline", siete también lo consideran el programa más fácil, de los nueve restantes: seis consideran que el programa más fácil es "Nmetbl" y tres "Ntree". De trece personas que dicen que el programa en el que detectaron más faltas es "Nmetbl", seis también lo consideran el programa más fácil, de los siete restantes: tres consideran que el más fácil es "Cmdline" y cuatro "Ntree". Finalmente, de ocho personas que dicen que el programa en el que detectaron más faltas es "Ntree", cuatro también lo consideran el programa más fácil y los cuatro restantes consideran que el más fácil es "Nmetbl".

Para la iteración 3, de dieciocho personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron más faltas es “Cmdline”, doce también lo consideran el programa más fácil, de los seis restantes: dos consideran que el más fácil es “Nametbl” y cuatro “Ntree”. De quince personas que dicen que el programa en el que detectaron más faltas es “Nametbl”, seis también lo consideran el programa más fácil, de los nueve restantes: cuatro consideran que el más fácil es “Cmdline” y cinco “Ntree”. Finalmente, de doce personas que dicen que el programa en el que detectaron más faltas es “Ntree”, cinco también lo consideran el programa más fácil, de los siete restantes: cuatro consideran que el más fácil es “Cmdline” y tres “Nametbl”.

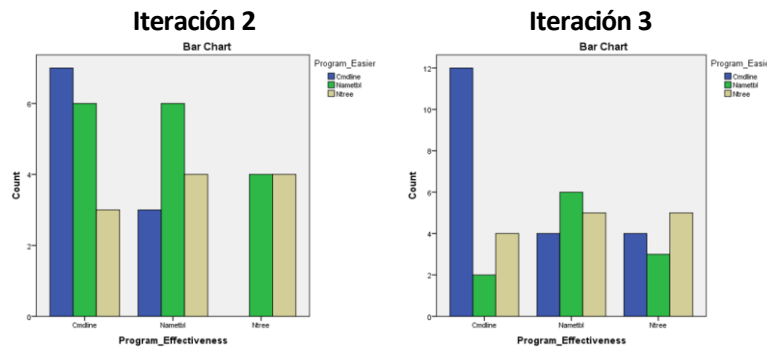


Figura 7.9. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.34 y Tabla 7.35 para las iteraciones 2 y 3, respectivamente. Donde se indica a que porcentaje de sujetos de cada uno de los programas.

			PP1			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	7	6	3	16
		% of Total	18,9%	16,2%	8,1%	43,2%
	Nametbl	Count	3	6	4	13
		% of Total	8,1%	16,2%	10,8%	35,1%
	Ntree	Count	0	4	4	8
		% of Total	0,0%	10,8%	10,8%	21,6%
Total		Count	10	16	11	37
		% of Total	27,0%	43,2%	29,7%	100,0%

Tabla 7.34. Tabla de Contingencia de la iteración 2

			PP1			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	12	2	4	18
		% of Total	26,7%	4,4%	8,9%	40,0%
	Nametbl	Count	4	6	5	15
		% of Total	8,9%	13,3%	11,1%	33,3%
	Ntree	Count	4	3	5	12
		% of Total	8,9%	6,7%	11,1%	26,7%
Total		Count	20	11	14	45
		% of Total	44,4%	24,4%	31,1%	100,0%

Tabla 7.35. Tabla de Contingencia de la iteración 3

7.2.1.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 2 es del 45.9%, es decir, casi la mitad de los sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 51.1%, es decir, poco más de la mitad de los sujetos están de acuerdo

En la Tabla 8.36, para la iteración 2, se puede observar que el acuerdo obtenido es debido al azar ($\text{sig} > 0.05$), por lo tanto, se puede concluir que no existe acuerdo.

Para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$), y dado que el valor Kappa muestra que el acuerdo es discreto (0,21 - 0,40), por lo tanto, se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 2	,189	,121	1,678	,093
Iteración 3	,257	,109	2,448	,014

Tabla 7.36. Índice de Concordancia Kappa

Debido a que en la iteración 2 no hay acuerdo y en la iteración 3 el acuerdo es discreto, se puede concluir que los sujetos no identifican el programa en el que han detectado más faltas con el programa más fácil.

7.2.2. ¿Con qué programa has detectado más faltas? / ¿Qué programa tiene las faltas más difíciles?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron más faltas y el programa que consideran era el más fácil. Esta pregunta aparece en la segunda y tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.2.1. Estudio del Acuerdo Después-Después

En la Figura 7.10, para la iteración 2, de dieciséis personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron más faltas es "Cmdline", dos también lo consideran el programa con faltas más difíciles, de los catorce restantes: seis dicen que el programa con faltas más difíciles es "Nmetbl" y ocho "Ntree". De trece personas que dicen que el programa en el que detectaron más faltas es "Nmetbl", uno también lo considera el programa con faltas más difíciles, de los doce restantes: seis dicen que el programa más fácil es "Cmdline" y seis "Ntree". Finalmente, de ocho personas que dicen que el programa en el que detectaron más faltas es "Ntree", ninguna lo considera el programa con faltas más difíciles, por lo tanto los ocho sujetos se distribuyen de la siguiente forma: cinco dicen que el programa con faltas más difíciles es "Cmdline" y tres "Nmetbl".

Para la iteración 3, de dieciocho personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron más faltas es "Cmdline", uno también lo considera el programa con

faltas más difíciles, de los diecisiete restantes: diez consideran que el programa con faltas más difíciles es “Nmetbl” y siete “Ntree”. De quince personas que dicen que el programa en el que detectaron más faltas es “Nmetbl”, dos también lo consideran el programa con faltas más difíciles, de los trece restantes: ocho consideran que el programa más fácil es “Cmdline” y cinco “Ntree”. Finalmente, de doce personas que dicen que el programa en el que detectaron más faltas es “Ntree”, dos también lo consideran el programa con faltas más difíciles, de los diez restantes: siete consideran que el programa con faltas más difíciles es “Cmdline” y tres “Nmetbl”.

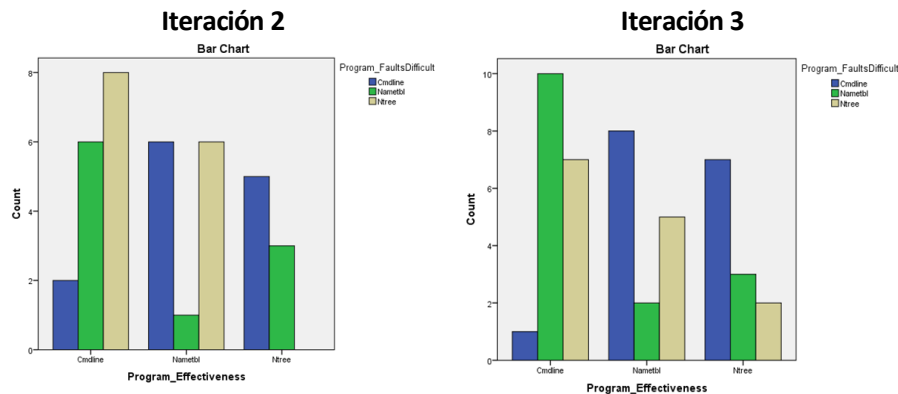


Figura 7.10. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.37 y en la Tabla 7.38 para las iteraciones 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP2			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	2	6	8	16
		% of Total	5,4%	16,2%	21,6%	43,2%
	Nametbl	Count	6	1	6	13
		% of Total	16,2%	2,7%	16,2%	35,1%
	Ntree	Count	5	3	0	8
		% of Total	13,5%	8,1%	0,0%	21,6%
Total		Count	13	10	14	37
		% of Total	35,1%	25,0%	37,8%	100,0%

Tabla 7.37. Tabla de Contingencia de la iteración 2

			PP2			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	1	10	7	18
		% of Total	2,2%	22,2%	15,6%	40,0%
	Nametbl	Count	8	2	5	15
		% of Total	17,8%	4,4%	11,1%	33,3%
	Ntree	Count	7	3	2	12
		% of Total	15,6%	6,7%	4,4%	26,7%
Total		Count	16	15	14	45
		% of Total	35,6%	33,3%	31,1%	100,0%

Tabla 7.38. Tabla de Contingencia de la iteración 3

7.2.2.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 2 es del 8.1%, es decir, muy pocos sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 11%, es decir, muy pocos sujetos están de acuerdo. Se esperaban estos resultados, puesto que las preguntas son contradictorias.

En la Tabla 8.39, para las Iteraciones 2 y 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$), y dado que el valor Kappa es negativo, se va a interpretar este valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 2	-,369	,069	-3,287	,001
Iteración 3	-,339	,075	-3,221	,001

Tabla 7.39. Índice de Concordancia Kappa

Debido a que en la iteración 2 no hay acuerdo y en la iteración 3 el acuerdo es discreto, se puede concluir que los sujetos no identifican el programa en el que han detectado más faltas con el programa más fácil.

7.2.3. ¿Con qué programa has detectado más faltas? / ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron más faltas y el programa que consideran que han entendido mejor. Esta pregunta aparece en la segunda y tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.3.1. Estudio de Acuerdo Después-Después

En la Figura 7.11, para la iteración 2, de dieciséis personas que dicen que el programa en el que detectaron más faltas es "Cmdline", seis también lo consideran el programa que han entendido mejor, de los diez restantes: seis consideran que el programa que han entendido mejor es "Nmetbl" y cuatro "Ntree". De trece personas que dicen que el programa en el que detectaron más faltas es "Nmetbl", ocho también lo consideran el programa que han entendido mejor, de los cinco restantes: dos consideran que el programa que han entendido mejor "Cmdline" y tres "Ntree". Finalmente, de ocho personas que dicen que el programa en el que detectaron más faltas es "Ntree", seis lo consideran el programa con faltas más difíciles y para los dos restantes el programa que han entendido mejor es "Nmetbl".

Para la iteración 3, de dieciocho personas que dicen que el programa en el que detectaron más faltas es "Cmdline", doce también lo considera el programa que han entendido mejor, de los seis restantes: uno consideran que el programa que han entendido mejor es "Nmetbl" y cinco "Ntree".

De quince personas que dicen que el programa en el que detectaron más faltas es “Nametbl”, siete también lo consideran el programa que han entendido mejor, de los ocho restantes: tres consideran que el programa que han entendido mejor es “Cmdline” y cinco “Ntree”. Finalmente, de doce personas que dicen que el programa en el que detectaron más faltas es “Ntree”, siete también lo consideran el programa que han entendido mejor, de los cinco restantes: dos consideran que el programa que han entendido mejor es “Cmdline” y tres “Nametbl”.

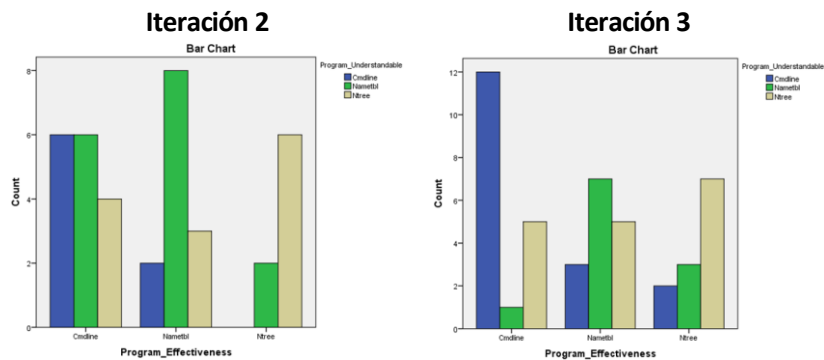


Figura 7.11. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.40 y Tabla 7.41 para las iteraciones 2 y 3, respectivamente. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP3			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	6	6	4	16
		% of Total	16,2%	16,2%	10,8%	43,2%
	Nametbl	Count	2	8	3	13
		% of Total	5,4%	21,6%	8,1%	35,1%
	Ntree	Count	0	2	6	8
		% of Total	0,0%	5,4%	16,2%	21,6%
Total		Count	8	16	13	37
		% of Total	21,6%	43,2%	35,1%	100,0%

Tabla 7.40. Tabla de Contingencia de la iteración 2

			PP3			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	12	1	5	18
		% of Total	26,7%	2,2%	11,1%	40,0%
	Nametbl	Count	3	7	5	15
		% of Total	6,7%	15,6%	11,1%	33,3%
	Ntree	Count	2	3	7	12
		% of Total	4,4%	6,7%	15,6%	26,7%
Total		Count	17	11	17	45
		% of Total	37,8%	24,4%	37,8%	100,0%

Tabla 7.41. Tabla de Contingencia de la iteración 3

7.2.3.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 2 es del 54%, es decir, más de la mitad de sujetos están de acuerdo. Finalmente para la iteración 3 el porcentaje es del 57.9%, es decir, más de la mitad sujetos están de acuerdo.

En la Tabla 8.42, para las Iteraciones 2 y 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el valor Kappa muestra un acuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 2	,323	,116	2,967	,003
Iteración 3	,367	,108	3,537	,000

Tabla 7.42. Índice de Concordancia Kappa

Debido a que en las dos iteraciones hay un acuerdo discreto, se puede concluir que parece que los sujetos identifican el programa en el que han detectado más faltas con el programa que han entendido mejor. La relación existente es baja.

7.2.4. ¿Con qué programa has detectado más faltas? / ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron más faltas y el programa que consideran que tiene las faltas más fáciles. Esta pregunta aparece solo en tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.4.1. Estudio del Acuerdo Después-Después

En la Figura 7.12, se observa que de dieciocho personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron más faltas es "Cmdline", trece también lo consideran el programa que tiene las faltas más fáciles, de los cinco restantes: uno considera que el programa que tiene las faltas más fáciles es "Nmetbl" y cuatro "Ntree". De quince personas que dicen que el programa en el que detectaron más faltas es "Nmetbl", diez también lo consideran el programa que tiene las faltas más fáciles, de los cinco restantes: tres consideran que el programa que tiene las faltas más fáciles es "Cmdline" y dos "Ntree". Finalmente, de doce personas que dicen que el programa en el que detectaron más faltas es "Ntree", tres lo consideran el programa que tiene las faltas más fáciles, de los nueve restantes: cuatro consideran que el programa que tiene las faltas más fáciles es "Cmdline" y cinco "Nmetbl".

Estos resultados se pueden contrastar en la Tabla 7.43. Donde además se indica a qué porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

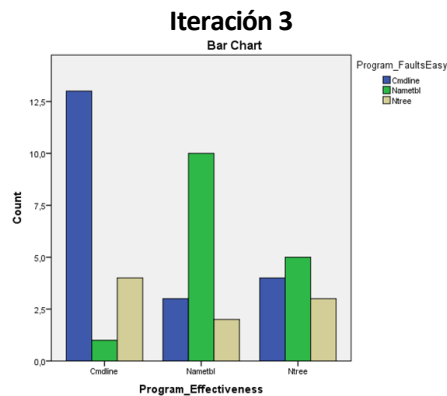


Figura 7.12. Acuerdo entre las respuestas

			PP1			Total
			Cmdline	Nametbl	Ntree	
IP1	Cmdline	Count	13	1	4	18
		% of Total	28,9%	2,2%	8,9%	40,0%
	Nametbl	Count	3	10	2	15
		% of Total	6,7%	22,2%	4,4%	33,3%
	Ntree	Count	4	5	3	12
		% of Total	8,9%	11,1%	6,7%	26,7%
Total		Count	20	16	9	45
		% of Total	44,4%	35,6%	20,0%	100,0%

Tabla 7.43. Tabla de Contingencia de la iteración 3

7.2.4.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 57.8%, es decir, más de la mitad de sujetos están de acuerdo.

En la Tabla 8.44, para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el valor Kappa muestra un acuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	,351	,107	3,299	,001

Tabla 7.44. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un acuerdo discreto, se puede concluir que parece que los sujetos identifican el programa en el que han detectado más faltas con el programa que tiene las faltas más fáciles. La relación existente es baja.

7.2.5. ¿Con qué programa has detectado menos faltas? / ¿Qué programa es más fácil?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron menos faltas y el programa que consideran era el más fácil. Esta

pregunta solo aparece en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.5.1. Estudio del Acuerdo Después-Después

En la Figura 7.13, se observa que de catorce personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron menos faltas es “Cmdline”, tres también lo consideran el programa más fácil, de los once restantes: cuatro consideran que el programa más fácil es “Nametbl” y siete “Ntree”. De quince personas que dicen que el programa en el que detectaron menos faltas es “Nametbl”, tres también lo consideran el programa más fácil, de los doce restantes: seis consideran que el más fácil es “Cmdline” y seis “Ntree”. Finalmente, de dieciséis personas que dicen que el programa en el que detectaron menos faltas es “Ntree”, uno también lo considera el programa más fácil, de los quince restantes: once consideran que el más fácil es “Cmdline” y cuatro “Nametbl”.

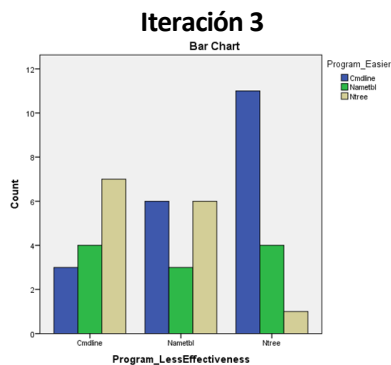


Figura 7.13. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.45. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP1			Total
			Cmdline	Nametbl	Ntree	
IP2	Cmdline	Count	3	4	7	14
		% of Total	6,7%	8,9%	15,6%	31,1%
	Nametbl	Count	6	3	6	15
		% of Total	13,3%	6,7%	13,3%	33,3%
	Ntree	Count	11	4	1	16
		% of Total	24,4%	8,9%	2,2%	35,6%
Total		Count	20	11	14	45
		% of Total	44,4%	24,4%	31,1%	100,0%

Tabla 7.45. Tabla de Contingencia de la iteración 3

7.2.5.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 15.6%, es decir, muy pocos sujetos están de acuerdo. Este resultado se esperaba puesto que las preguntas son contradictorias.

En la Tabla 7.46, para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor Kappa es negativo, se va a interpretar este valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	-,261	,084	-2,534	,011

Tabla 7.46. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un desacuerdo discreto, se puede concluir que, como era de esperar, parece que los sujetos no identifican el programa en el que han detectado menos faltas con el programa más fácil.

7.2.6. ¿Con qué programa has detectado menos faltas? / ¿Qué programa tiene las faltas más difíciles?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron menos faltas y el programa que consideran tiene las faltas más difíciles. Esta pregunta aparece solo en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.6.1. Estudio del Acuerdo Después-Después

En la Figura 7.14, se observa que de quince personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron menos faltas es "Cmdline", once también lo consideran el programa con faltas más difíciles, de los cuatro restantes: dos consideran que el programa con faltas más difíciles es "Nmetbl" y dos "Ntree". De quince personas que dicen que el programa en el que detectaron menos faltas es "Nmetbl", diez también lo consideran el programa con faltas más difíciles, de los cinco restantes: cuatro consideran que el programa más fácil es "Cmdline" y uno "Ntree". Finalmente, de dieciséis personas que dicen que el programa en el que detectaron menos faltas es "Ntree", once lo consideran el programa con faltas más difíciles, de los cinco restantes: dos consideran que el programa con faltas más difíciles es "Cmdline" y tres "Nmetbl".

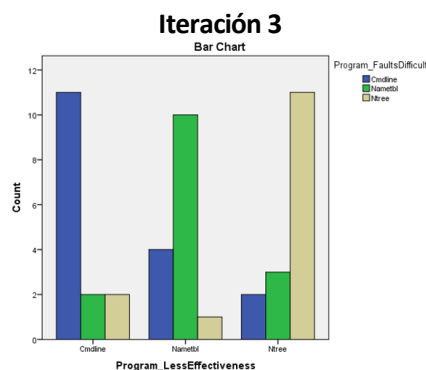


Figura 7.14. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.47. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP2			Total
			Cmdline	Nametbl	Ntree	
IP2	Cmdline	Count	11	2	2	15
		% of Total	23,9%	4,3%	4,3%	32,6%
	Nametbl	Count	4	10	1	15
		% of Total	8,7%	21,7%	2,2%	32,6%
	Ntree	Count	2	3	11	16
		% of Total	4,3%	6,5%	23,9%	34,8%
Total		Count	17	15	14	46
		% of Total	37,0%	32,6%	30,4%	100,0%

Tabla 7.47. Tabla de Contingencia de la iteración 3

7.2.6.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 69.5%, es decir, la mayoría de los sujetos están de acuerdo.

En la Tabla 8.48, para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y el valor Kappa muestra que existe un acuerdo moderado ($>0,41 - 0,60$), por lo tanto, se puede concluir que existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	,544	,101	5,234	,000

Tabla 7.48. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un acuerdo moderado, se puede concluir que los sujetos identifican el programa en el que han detectado menos faltas con el programa que tiene las faltas más difíciles. La relación existente es intermedia.

7.2.7. ¿Con qué programa has detectado menos faltas? / ¿Qué programa has entendido mejor?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron menos faltas y el programa que consideran que han entendido mejor. Esta pregunta solo aparece en la tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.7.1. Estudio del Acuerdo Después-Después

En la Figura 7.15, para la iteración 3, de catorce personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron menos faltas es "Cmdline", dos también lo consideran el programa que han entendido mejor, de los doce restantes: seis consideran que el

programa que han entendido mejor es “Nametbl” y seis “Ntree”. De quince personas que dicen que el programa en el que detectaron menos faltas es “Nametbl”, uno también lo considera el programa que han entendido mejor, de los catorce restantes: seis consideran que el programa que han entendido mejor “Cmdline” y ocho “Ntree”. Finalmente, de dieciséis personas que dicen que el programa en el que detectaron menos faltas es “Ntree”, tres lo consideran el programa con faltas más difíciles, de los trece restantes: nueve consideran que el programa que han entendido mejor es “Cmdline” y cuatro “Nametbl”.

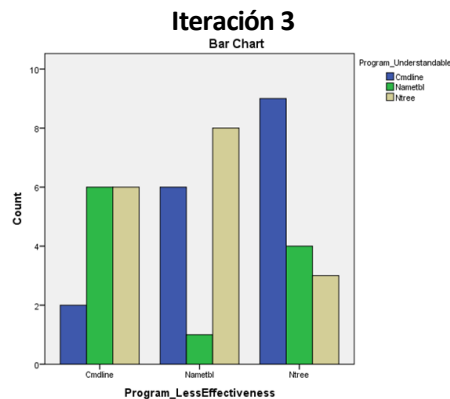


Figura 7.15. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.49. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP3			Total
			Cmdline	Nametbl	Ntree	
IP2	Cmdline	Count	2	6	6	14
		% of Total	4,4%	13,3%	13,3%	31,1%
	Nametbl	Count	6	1	8	15
		% of Total	13,3%	2,2%	17,8%	33,3%
	Ntree	Count	9	4	3	16
		% of Total	20,0%	8,9%	6,7%	35,6%
Total		Count	17	11	17	45
		% of Total	37,8%	24,4%	37,8%	100,0%

Tabla 7.49 Tabla de Contingencia de la iteración 3

7.2.7.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 13.3%, es decir, muy pocos sujetos están de acuerdo. Se esperaba este resultado, puesto que las preguntas son contradictorias.

En la Tabla 8.92, para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor Kappa es negativo, se va a interpretar este valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	-,300	,075	-2,874	,004

Tabla 7.50. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un desacuerdo discreto, se puede concluir que, como era de esperar, parece que los sujetos no identifican el programa en el que han detectado menos faltas con el programa que han entendido mejor.

7.2.8. ¿Con qué programa has detectado menos faltas? / ¿Qué programa tiene las faltas más fáciles?

En esta sección se va a evaluar la relación existente entre el programa en el que los sujetos consideran que detectaron menos faltas y el programa que consideran que tiene las faltas más fáciles. Esta pregunta aparece solo en tercera iteración. Dado que ambas preguntas solo se realizaron después del estudio empírico, se va a analizar la combinación: Después-Después.

7.2.8.1. Estudio del Acuerdo Después-Después

En la Figura 7.16, para la iteración 3, de quince personas que dicen, después de aplicar las técnicas, que el programa en el que detectaron menos faltas es "Cmdline", tres también lo consideran el programa que tiene las faltas más fáciles, de los doce restantes: diez consideran que el programa que tiene las faltas más fáciles es "Ntree" y dos "Cmdline". De quince personas que dicen que el programa en el que detectaron menos faltas es "Ntree", uno también lo considera el programa que tiene las faltas más fáciles, de los catorce restantes: siete consideran que el programa que tiene las faltas más fáciles es "Cmdline" y siete "Ntree". Finalmente, de dieciséis personas que dicen que el programa en el que detectaron más fácil es "Ntree", ninguna lo consideran el programa que tiene las faltas más fáciles, por lo tanto los dieciséis sujetos se distribuyen así: diez consideran que el programa que tiene las faltas más fáciles es "Cmdline" y seis "Ntree".

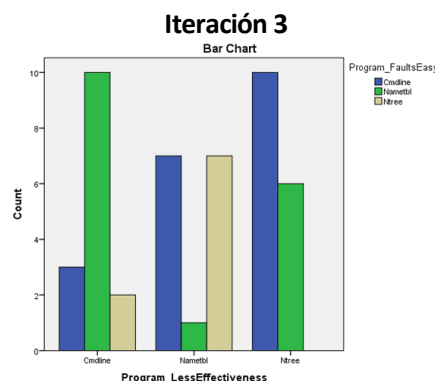


Figura 7.16. Acuerdo entre las respuestas

Estos resultados se pueden contrastar en la Tabla 7.96. Donde además se indica a que porcentaje de sujetos corresponden las respuestas de cada uno de los programas.

			PP1			Total
			Cmdline	Nametbl	Ntree	
IP2	Cmdline	Count	3	10	2	15
		% of Total	6,5%	21,7%	4,3%	32,6%
	Nametbl	Count	7	1	7	15
		% of Total	15,2%	2,2%	15,2%	32,6%
	Ntree	Count	10	6	0	16
		% of Total	21,7%	13,0%	0,0%	34,8%
Total		Count	20	17	9	46
		% of Total	43,5%	37,0%	19,6%	100,0%

Tabla 7.51. Tabla de Contingencia de la iteración 3

7.2.8.1.1. Medida de Acuerdo

En las tablas de contingencia, se puede observar que el porcentaje de acuerdo para la iteración 3 es del 8.7%, es decir, muy pocos sujetos están de acuerdo. Se esperaba este resultado, puesto que las preguntas son contradictorias.

En la Tabla 8.97, para la iteración 3, se puede observar que el acuerdo obtenido no es debido al azar ($\text{sig} \leq 0.05$) y dado que el valor Kappa es negativo, se va a interpretar este valor como desacuerdo, es decir, existe un desacuerdo discreto (0,21 - 0,40), por lo tanto, se puede concluir que no existe acuerdo.

Iteración	Valor	Error típ. asint.	T aproximada	Sig. aproximada
Iteración 3	-,363	,062	-3,596	,000

Tabla 7.52. Índice de Concordancia Kappa

Debido a que en la iteración 3 hay un desacuerdo discreto, se puede concluir que, como era de esperar, parece que los sujetos no identifican el programa en el que han detectado menos faltas con el programa que tiene las faltas más fáciles.

CAPÍTULO 8. DISCUSIÓN

8.1. Resultados sobre Intuición

En esta sección se va a discutir sobre los resultados obtenidos en el capítulo 3, primero se explican los resultados sobre las técnica, luego los resultado sobre los programas y finalmente los resultados de las consistencias entre las preguntas.

8.1.1. Resultados sobre las Técnicas

Todos los resultados mostrados a continuación, se pueden contrastar en la Tabla 8.1, donde se muestra la distribución de la técnica y en la Tabla 8.2 donde se muestra el acuerdo entre las preguntas que aparecen antes y después de aplicar las técnicas.

Pregunta		Iteración 1		Iteración 2		Iteración 3	
		X ²	Valor	X ²	Valor	X ²	Valor
IT1	A	,260	R=E=F	,424	R=E=F	,071	R=E=F
	D	,260	R=E=F	,161	R=E=F	,013	(R=F)>E
IT2	D	,015	F>E>R	,004	F>(R=E)	,005	(R=F)>E
IP1	D	No hay resultados		,266	cm=na=nt	,549	cm=na=nt
IP2	D	No hay resultados		No hay resultados		,978	cm=na=nt

Tabla 8.1. Resumen Distribución de Intuiciones

Pregunta	Iteración 1				Iteración 2				Iteración 3			
	Kappa*	McNemar-Bowker*	Valor	Stuart-Maxwell*	Kappa*	McNemar-Bowker*	Valor	Stuart-Maxwell*	Kappa*	McNemar-Bowker*	Valor	Stuart-Maxwell*
IT1	,074	Azar	1	1	,712	Azar	,627	,454	,282	Azar	,011	,004

*Los valores mostrados son la significancia obtenida (sig)

Tabla 8.2. Resumen de Cambios de Opinión

¿Qué técnica detecta más faltas? (IT1)

En la iteración 1, se observa que antes de aplicar las técnicas, no hay una “intuición clara”, o dicho de otra manera, los sujetos intuyen que las tres técnicas son igual de efectivas. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, sin embargo, después de este cambio tampoco hay una “intuición clara”. Los cambios más relevantes son de la técnica “Estructural” hacia “Funcional” y de “Funcional” hacia “Estructural” pocos sujetos cambian de “Revisión” a “Funcional” y de “Revisión” a “Funcional”. A pesar de esto los cambios no tienen dirección. Los cambios son de la siguiente manera: $R \rightarrow E = E \rightarrow R$, $E \rightarrow F = F \rightarrow E$, $R \rightarrow F = F \rightarrow R$, puesto que las diferencias de cambios es igual a 0 en los tres casos.

En la iteración 2, se observa que antes de aplicar las técnicas, no hay una “intuición clara”, o dicho de otra manera, los sujetos intuyen que las tres técnicas son igual de efectivas. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, sin embargo, después de este cambio tampoco hay una “intuición clara”. Los cambios más relevantes son de la técnica “Estructural” hacia “Funcional” y de “Revisión” hacia “Estructural” y “Funcional” indistintamente, finalmente también hay cambios de “Funcional” hacia “Revisión” y “Estructural”. A pesar de esto los cambios no tienen dirección. Los cambios son de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, los cambios de $E \rightarrow F = F \rightarrow E$ y hay muchos más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 1, 0 y 4 respectivamente.

En la iteración 3, se observa que antes de aplicar las técnicas, no hay una “intuición clara”, o dicho de otra manera, los sujetos intuyen que las tres técnicas son igual de efectivas. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, y después de este cambio los sujetos intuyen que las técnicas “Revisión” y “Funcional” son más efectivas que “Estructural”. Los cambios más relevantes son de la técnica “Estructural” hacia “Revisión” y “Funcional” indistintamente y de “Revisión” hacia “Funcional” y de “Estructural” hacia “Funcional”, finalmente hay pocos cambios de “Revisión” hacia “Estructural” y de “Funcional” hacia “Revisión”. Además los cambios tienen dirección y los mismos se producen de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, hay más cambios de $E \rightarrow F$ que de $F \rightarrow E$ y finalmente hay más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 5, 5 y 6, respectivamente.

En resumen, para las iteraciones 1 y 2 se obtienen resultados consistentes: no hay una tendencia clara en las intuiciones de los sujetos. Asimismo, estos tienden a cambiar de opinión tras la aplicación de las técnicas, sin embargo, esto se produce de manera indistinta (no se identifica ningún patrón). Finalmente tras la aplicación de las técnicas, se vuelve a observar la inexistencia de una tendencia relativa a las intuiciones. En la iteración 3, aunque inicialmente no hay tendencia en las intuiciones, tras aplicar las técnicas los sujetos parecen decantarse por las técnicas “Revisión” y “Funcional”, que intuyen que son más efectivas que “Estructural”.

¿Qué técnica has aplicado mejor? (IT2)

Con las repuestas proporcionadas después de aplicar todas las técnicas, se puede observar que, en las tres iteraciones, los sujetos intuyen que la técnica que han aplicado mejor es “Funcional”,

sin embargo en la iteración 1, también se puede aceptar que los sujetos han intuitido “Estructural” y en la iteración 3 “Revisión”.

8.1.2. Resultados sobre los Programas

¿Con qué programa has detectado más faltas? (IP1)

Con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, no hay una tendencia de intuición, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa más efectivo.

¿Con qué programa has detectado menos faltas? (IP2)

De forma similar a lo que ocurre en la pregunta anterior, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que no hay una tendencia de intuición, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa menos efectivo.

8.1.3. Consistencias entre las Preguntas

En la Tabla 8.3 se observan los valores para cada par de preguntas.

¿Qué técnica detecta más falta? (IT1) / ¿Qué técnica has aplicado mejor? (IT2)

Los sujetos de la iteración 1, no están de acuerdo, sin embargo para las Iteraciones 2 y 3, los sujetos están identificando la técnica más efectiva, con la que han aplicado mejor. En conclusión, dado que aunque en dos iteraciones hay acuerdo, en la iteración 3 el acuerdo es bajo, parece que existe cierta tendencia a contestar lo mismo. Es decir, existe poca relación entre la técnica más efectiva y la técnica que han aplicado mejor.

¿Con qué programa has detectado más falta? (IP1) / ¿Con qué programa has detectado menos falta? (IP2)

Los sujetos, no están de acuerdo, es decir, los sujetos no están identificando el programa más efectivo, con el programa menos efectivo. Se esperaba que no hubiera acuerdo, dado que las preguntas son contradictorias.

Pregunta	Iteración 1		Iteración 2			Iteración 3		
	Kappa		Kappa			Kappa		
IT1/IT2	,348	Azar	,000	,561	Moderado	,001	,384	Discreto
IP1/IP2	No hay resultados		No hay resultados			,000	-,427	Moderado

Tabla 8.3. Tabla Resumen: Consistencia Intuición

8.1.4. Resultados Globales

En definitiva, con respecto a las técnicas, no parece que haya una intuición clara sobre la más efectiva, ni antes ni después de su aplicación, a pesar de que los sujetos tienden a cambiar de opinión. Sin embargo, los sujetos parecen coincidir en que aplican mejor la técnica “Funcional”

obteniéndose resultados contradictorios para “Revisión” y “Estructural”. Por otra parte, en cierta manera, los sujetos tienden a identificar la técnica más efectiva con la técnica que han aplicado mejor, aunque la relación existente es baja.

Con respecto a los programas, no parece que haya una intuición clara sobre el más efectivo, ni tampoco con el menos efectivo. Por otra parte, los sujetos no tienden a identificar el programa más efectivo con el menos efectivo.

8.2. Resultados sobre Preferencias

En esta sección se va a discutir sobre los resultados obtenidos en el capítulo 5, primero se explican los resultados sobre las técnicas, luego los resultados sobre los programas y finalmente los resultados de las consistencias entre las preguntas.

8.2.1. Resultados sobre las Técnicas

Todos los resultados mostrados a continuación, se pueden contrastar en la Tabla 8.4, donde se muestra la distribución de la técnica y en la Tabla 8.5 donde se muestra el acuerdo entre las preguntas que aparecen antes y después de aplicar las técnicas.

Pregunta		Iteración 1		Iteración 2		Iteración 3	
		X ²	Valor	X ²	Valor	X ²	Valor
PT1	A	,032	E>F>R	,000	F>(R=E)	,223	R=E=F
	D	,000	F>(E>R)	,000	F>(R=E)	,000	F>(R>E)
PT2	D	,223	R=E=F	,004	F>R>E	,157	R=E=F
PT3	A	,074	R=E=F	,000	F>(R=E)	,004	F>(R=E)
	D	,001	F>(R=E)	,000	F>(R=E)	,000	F>(R=E)
PP1	D	,156	cm=na=nt	,479	cm=na=nt	,247	cm=na=nt
PP2	D	,178	cm=na=nt	,832	cm=na=nt	,859	cm=na=nt
PP3	D	No hay resultados		,378	cm=na=nt	,449	cm=na=nt
PP4	D	No hay resultados		No hay resultados		,121	cm=na=nt

Tabla 8.4. Resumen Distribución de Preferencias

¿Qué técnica te gusta más? (PT1)

En la iteración 1, se observa que antes de aplicar las técnicas, la técnica que gusta más a los sujetos es “Estructural”, aunque también se puede aceptar “Funcional”. La aplicación de la técnica no ha hecho que los sujetos cambien de opinión, puesto que hay un acuerdo moderado. Sin embargo, los pocos cambios que se producen han hecho que ahora la técnica que más gusta a los sujetos es “Funcional”, aunque también se puede aceptar “Estructural”. Los cambios se producen de la técnica “Estructural” hacia “Funcional” y muy pocos sujetos cambian de “Revisión” hacia “Funcional”. Los cambios tienen dirección y son de la siguiente manera: los cambios de $R \rightarrow E = E \rightarrow R$, hay más cambios de $E \rightarrow F$ que de $F \rightarrow E$ y hay muchos más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 0, 1 y 7 respectivamente.

En la iteración 2, se observa que antes de aplicar las técnicas, la técnica preferida de los sujetos es “Funcional”. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, sin

embargo, después del cambio los sujetos siguen prefiriendo “Funcional”. Los cambios más relevantes son de la técnica “Revisión” hacia “Funcional” y de “Funcional” hacia “Revisión”; de “Estructural” hacia “Revisión” y “Funcional” indistintamente, finalmente hay pocos cambios de “Funcional” hacia “Estructural”. A pesar de esto los cambios no tienen dirección, puesto que se producen de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, hay más cambios de $E \rightarrow F$ que de $F \rightarrow E$, finalmente hay más cambios de $F \rightarrow R$ que de $R \rightarrow F$, puesto que las diferencias de cambios es 1, en los tres casos.

	Iteración 1				Iteración 2				Iteración 3								
Pregunta	Kappa*			McNemar-Bowker*	Valor	Stuart-Maxwell*	Kappa *		McNemar-Bowker*	Valor	Stuart-Maxwell*	Kappa*			McNemar-Bowker*	Valor	Stuart-Maxwell*
PT1	,001	,475	Moderado	,018	R-E: 0	,018	,102	Azar	,695	R-E: -1	1	,016	,256	Discreto	,044	R-E: -2	,018
					E-F: 1					E-F: 1						E-F: -1	
					R-F: 7					R-F: -1						R-F: 6	
PT3	,000	,502	Moderado	,362	R-E: 1	,204	,102	Azar	,695	R-E: -1	1	,112	Azar	,286	R-E: -2	,152	
					E-F: 2					E-F: 1					E-F: 0		
					R-F: 1					R-F: -1					R-F: 5		

*Los valores mostrados son la significancia obtenida (sig)

Tabla 8.5. Resumen de Cambios de Opinión

En la iteración 3, antes de aplicar las técnicas, no hay una “preferencia clara” o dicho de otra manera, a los sujetos les gustan las tres técnicas por igual. La aplicación de la técnica no ha hecho que los sujetos cambien de opinión, puesto que hay un acuerdo discreto. Sin embargo, los cambios que se producen han hecho que la técnica preferida de los sujetos es “Funcional” y también se puede aceptar “Revisión”. Los cambios se producen de la técnica “Revisión” hacia “Funcional”, de “Estructural” hacia “Funcional” y de “Funcional” hacia “Revisión” y muy pocos sujetos cambian de “Estructural” hacia “Revisión”. Los cambios tienen dirección y son de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, hay más cambios de $F \rightarrow E$ que de $E \rightarrow F$, finalmente hay más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 2, 1 y 6, respectivamente.

En resumen, para las tres iteraciones se obtienen resultados contradictorios antes de aplicar las técnicas: en la iteración los sujetos prefieren “Estructural”, en la iteración 2 prefieren “Funcional” y en la tercera iteración no hay una tendencia clara en las preferencias de los sujetos. Asimismo, en la iteración 3 se puede decir que existe cambio, puesto que el acuerdo es bajo, por lo tanto, en las iteraciones 2 y 3, los sujetos tienden a cambiar de opinión tras la aplicación de las técnicas, y en la iteración 1 hay más acuerdo, aunque tampoco es muy alto. Por lo tanto, finalmente tras la aplicación de las técnicas, los sujetos consideran que su técnica preferida es “Funcional”.

¿Qué técnica te ha parecido más fácil de entender? (PT2)

Con las repuestas proporcionadas después de aplicar las técnicas, en la iteración 2a los sujetos les ha parecido más fácil de entender “Funcional” también se puede aceptar “Revisión”, sin embargo en las iteraciones 1 y 3 no hay una “preferencia clara” o dicho de otra manera, a los sujetos les ha parecido que las tres técnicas son igual de fáciles de entender.

¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

En la iteración 1, se observa que antes de aplicar las técnicas no hay una “preferencia clara”, o dicho de otra manera, los sujetos consideran que las tres técnicas son igual de fáciles de aplicar. La aplicación de la técnica no ha hecho que los sujetos cambien de opinión, puesto que hay un acuerdo moderado. Sin embargo, los pocos cambios han hecho que los sujetos consideren que la técnica más fácil de aplicar es “Funcional”. Los cambios que se producen son de la técnica “Revisión” hacia “Funcional” y “Estructural” indistintamente, de “Estructural” hacia “Funcional” y muy pocos sujetos cambian de “Funcional” hacia “Estructural”. Los cambios no tienen dirección y son de la siguiente manera: hay más cambios de $R \rightarrow E$ que de $E \rightarrow R$, hay más cambios de $E \rightarrow F$ que de $F \rightarrow E$ y hay más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 1, 2 y 1 respectivamente.

En la iteración 2, se observa que antes de aplicar las técnicas, la técnica que los sujetos consideran la más fácil de aplicar es “Funcional”. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, sin embargo, después del cambio los sujetos siguen creyendo que la técnica más fácil de aplicar es “Funcional”. Los cambios que se producen son de la técnica “Revisión” hacia “Funcional”, de “Estructural” hacia “Revisión” y “Funcional” indistintamente, de “Funcional” hacia “Revisión” y muy pocos sujetos cambian de “Funcional” hacia “Estructural”. A pesar de esto los cambios no tienen dirección. Los cambios son de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, hay más cambios de $E \rightarrow F$ que de $F \rightarrow E$, finalmente hay más cambios de $F \rightarrow R$ que de $R \rightarrow F$, puesto que las diferencias de cambios es 1, en los tres casos.

En la iteración 3, se observa que antes de aplicar las técnicas, la técnica que los sujetos consideran la más fácil de aplicar es “Funcional”. La aplicación de la técnica ha hecho que los sujetos cambien de opinión, sin embargo, después del cambio los sujetos siguen creyendo que la técnica más fácil de aplicar es “Funcional”. Los cambios que se producen son de la técnica “Revisión” hacia “Funcional”, de “Estructural” hacia “Funcional”, de “Funcional” hacia “Revisión” y “Estructural” indistintamente, y muy pocos sujetos cambian de “Revisión” hacia “Estructural” y de “Estructural” hacia “Revisión”. A pesar de esto los cambios no tienen dirección. Los cambios son de la siguiente manera: hay más cambios de $E \rightarrow R$ que de $R \rightarrow E$, los cambios de $E \rightarrow F = F \rightarrow E$, finalmente hay más cambios de $R \rightarrow F$ que de $F \rightarrow R$, puesto que las diferencias de cambios son 2, 0 y 5, respectivamente.

En resumen, para las iteraciones 2 y 3 se obtienen resultados consistentes: antes de aplicar las técnicas los sujetos consideran que la técnica más fácil de aplicar es “Funcional”. Asimismo, estos tienden a cambiar de opinión tras la aplicación de las técnicas, sin embargo, esto no hace que cambie su elección. Puesto que los sujetos, sigue creyendo que la técnica más fácil de aplicar es “Funcional”. En la iteración 1, aunque inicialmente no hay tendencia en las preferencias, aunque no hay cambios de opinión, tras aplicar las técnicas los sujetos se decantan por la técnica “Funcional”.

8.2.2. Resultados sobre los Programas

¿Qué programa es más fácil? (PP1)

Con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, no hay una tendencia de preferencia, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa más fácil.

¿Qué programa tiene las faltas más difíciles? (PP2)

De forma similar a lo que ocurre en la pregunta anterior, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, no hay una tendencia de preferencia, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa con las faltas más difíciles.

¿Qué programa has entendido mejor? (PP3)

De forma similar a lo que ocurre en la pregunta anterior, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, no hay una tendencia de preferencia, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa que han entendido mejor.

¿Qué programa tiene las faltas más fáciles? (PP4)

De forma similar a lo que ocurre en la pregunta anterior, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, no hay una tendencia de preferencia, en ninguna de las dos iteraciones. Dicho de otra manera, los sujetos se distribuyen uniformemente sobre el programa con las faltas más fáciles.

8.2.3. Consistencias entre las Preguntas

En la Tabla 8.5 se observan los valores para cada par de preguntas.

¿Qué técnica te gusta más? (PT1) / ¿Qué técnica te ha parecido más fácil de entender? (PT2)

En ninguna de las tres iteraciones, los sujetos están de acuerdo, por lo tanto, no existe tendencia a contestar lo mismo, es decir, los sujetos no identifican la técnica preferida con la más fácil de entender.

¿Qué técnica te gusta más? (PT1) / ¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

Tanto para en antes como para el después, en la iteración 1, no hay de acuerdo, sin embargo, en la iteración 2, es acuerdo es total y en la iteración 3 el acuerdo es discreto, por lo tanto, parece que existe tendencia a contestar lo mismo. Debido a que en dos iteraciones hay acuerdo y en una de las mimas el acuerdo es absoluto, se puede concluir que los sujetos identifican la técnica preferida con la más fácil de aplicar. Hay que tener en cuenta que en general, la relación es intermedia.

¿Qué técnica te ha parecido más fácil de entender? (PT2) / ¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

En ninguna de las tres iteraciones, los sujetos están de acuerdo, por lo tanto, no existe tendencia a contestar lo mismo, es decir, los sujetos no identifican la técnica más fácil de entender con la más fácil de aplicar.

		Iteración 1		Iteración 2			Iteración 3		
Pregunta		Kappa		Kappa			Kappa		
PT1/PT2		,783	Azar	,339	Azar		,108	Azar	
PT1/PT3	A	,318	Azar	,000	1	Absoluto	,014	,271	Discreto
	D	,444	Azar	,000	1	Absoluto	,003	,346	Discreto
PT2/PT3		,840	Azar	,339	Azar		,233	Azar	
PP1/PP2		,431	Azar	,003	-,332	Discreto	,000	-,373	Discreto
PP1/PP3		No hay resultado		,000	,677	Sustancial	,000	,593	Moderado
PP2/PP3		No hay resultado		,0003	-,333	Discreto	,000	-,400	Discreto
PP1/PP4		No hay resultado		No hay resultado			,000	,456	Moderado
PP2/PP4		No hay resultado		No hay resultado			,000	-,484	Moderado
PP3/PP4		No hay resultado		No hay resultado			,000	,502	Moderado

Tabla 8.6. Tabla Resumen: Consistencia Preferencias

¿Qué programa es más fácil? (PP1) / ¿Qué programa tiene las faltas más difíciles? (PP2)

En la iteración 1, no hay de acuerdo, sin embargo en las iteraciones 2 y 3 hay un desacuerdo discreto, por lo tanto, no existe tendencia a contestar lo mismo. Debido a que en dos iteraciones hay desacuerdo y en la iteración 1 se puede considerar el no acuerdo como desacuerdo, se puede concluir que, tal como era de esperar, los sujetos no identifican el programa más fácil con el programa que tiene las faltas más difíciles.

¿Qué programa es más fácil? (PP1) / ¿Qué programa has entendido mejor? (PP3)

En la iteración 2, hay un acuerdo sustancial y en la iteraciones 3 hay un acuerdo moderado, se puede concluir que, existe tendencia a contestar lo mismo, es decir, los sujetos identifican el programa más fácil con el programa que han entendido mejor. Hay que tener en cuenta que en general, la relación es intermedia.

¿Qué programa tiene las faltas más difíciles? (PP2) / ¿Qué programa has entendido mejor? (PP3)

En las iteraciones 2 y 3 hay un desacuerdo discreto, por lo tanto, no existe tendencia a contestar lo mismo, tal como era de esperar, los sujetos no identifican el programa que tiene las faltas más difíciles con el programa que han entendido mejor.

¿Qué programa es más fácil? (PP1) / ¿Qué programa tiene las faltas más fáciles? (PP4)

En la iteración 3 hay un acuerdo moderado, por lo tanto, existe tendencia a contestar lo mismo, es decir, los sujetos identifican el programa más fácil con el programa que tiene las faltas más fáciles. Hay que tener en cuenta que en general, la relación es intermedia.

¿Qué programa tiene las faltas más difíciles? (PP2) / ¿Qué programa tiene las faltas más fáciles? (PP4)

En la iteración 3 hay un desacuerdo moderado, por lo tanto, no existe tendencia a contestar lo mismo, tal como era de esperar, los sujetos no identifican el programa que tiene las faltas más difíciles con el programa que tiene las faltas más fáciles.

¿Qué programa has entendido mejor? (PP3) / ¿Qué programa tiene las faltas más fáciles? (PP4)

En la iteración 3 hay un acuerdo moderado, por lo tanto, existe tendencia a contestar lo mismo, los sujetos identifican el programa que han entendido mejor con el programa que tiene las faltas más fáciles. Hay que tener en cuenta que en general, la relación es intermedia.

8.2.4. Resultados Globales

En definitiva, con respecto a las técnicas, no parece que haya una decisión clara sobre la técnica preferida y la más fácil de aplicar, antes de aplicar las técnicas, además los sujetos tienden a cambiar de opinión, este cambio, conlleva a que después de aplicar las técnicas coinciden en que la técnica preferida y la más fácil de aplicar es “Funcional”. Sin embargo, los sujetos no tienen una elección clara sobre la técnica más fácil de entender. Por otra parte, los sujetos no identifican la técnica preferida con la más fácil de aplicar, ni tampoco identifican la técnica más fácil de entender con la más fácil de aplicar. Sin embargo, si identifican, tanto antes como después de aplicar las técnicas, la técnica preferida con la más fácil de aplicar, aunque la relación existente en general, es intermedia.

Con respecto a los programas, no parece que haya una preferencia clara sobre más fácil, ni el que tiene las faltas más difíciles, ni el que ellos han entendido mejor, ni tampoco el que tiene las faltas más fáciles. Por otra parte, los sujetos no tienden a identificar el programa más fácil con el que tiene las faltas más difíciles, ni tampoco el que tiene las faltas más difíciles con el que han entendido mejor, ni tampoco el que tiene las faltas más difíciles con el que tiene las faltas más fáciles. Sin embargo, si identifican el programa más fácil, tanto con el que han entendido mejor como con el que tiene las faltas más fáciles y también identifican el programa que han entendido mejor con el que tiene las faltas más fáciles. La relación existente es intermedia.

8.3. Acuerdo entre las Intuiciones y Preferencias

En esta sección se va a discutir sobre el acuerdo entre las preguntas de intuiciones y las preguntas de preferencias, para ello se muestra todas las combinaciones posibles de las preguntas de los dos tipos, comparando siempre una de pregunta de intuición frente a una de preferencia. Los resultados, de las técnicas y los programas se pueden contrastar en la Tabla 8.7.

		Iteración 1		Iteración 2		Iteración 3			
Pregunta		Kappa		Kappa		Kappa			
IT1/PT1	A	,214	Azar	,736	Azar		,041	,198	Insignificante
	D	,078	Azar	,004	,325	Discreto	,000	,404	Discreto
IT1/PT2		,386	Azar	,769	Azar		,784	Azar	
IT1/PT3	A	,075	Azar	,736	Azar		,113	Azar	
	D	,088	Azar	,004	,325	Discreto	,041	,206	Discreto
IT2/PT1		,738	Azar	,000	,652	Sustancial	,000	,573	Moderado
IT2/PT2		,422	Azar	,169	Azar		,159	Azar	
IT2/PT3		,004	,417	,000	,652	Sustancial	,000	,426	Moderado
IP1/PP1		No hay resultado		,093	Azar		,014	,257	Discreto
IP1/PP2		No hay resultado		,001	-,369	Discreto	,001	-,339	Discreto
IP1/PP3		No hay resultado		,003	,323	Discreto	,000	,367	Discreto
IP1/PP4		No hay resultado		No hay resultado			,001	,351	Discreto
IP2/PP1		No hay resultado		No hay resultado			,011	-,261	Discreto
IP2/PP2		No hay resultado		No hay resultado			,000	,544	Moderado
IP2/PP3		No hay resultado		No hay resultado			,004	-,300	Discreto
IP2/PP4		No hay resultado		No hay resultado			,000	-,363	Discreto

Tabla 8.7. Tabla Resumen: Consistencia Intuición vs Preferencias

8.3.1. Resultados sobre las Técnicas

¿Qué técnica detecta más faltas? (IT1) / ¿Qué técnica te gusta más? (PT1)

Antes de aplicar las técnicas, ni en la iteración 1 ni en la 2, existe acuerdo y en la iteración 3, hay un acuerdo insignificante, por lo tanto, parece que no existe tendencia a contestar lo mismo. Debido a que en dos iteraciones no hay acuerdo y en la otra es acuerdo es muy bajo, se puede concluir que los sujetos no identifican la técnica más efectiva con la técnica que les gusta más.

Después de aplicar las técnicas, en la iteración 1, no hay acuerdo, sin embargo, en la iteración 2 y 3, hay un acuerdo discreto, por lo tanto, parece que existe tendencia a contestar lo mismo. Debido a que en dos iteraciones hay acuerdo (aunque es bajo) y solo en una no hay acuerdo, se puede concluir que, debido a que los sujetos cambian de opinión, parece que después los sujetos identifican la técnica más efectiva con la técnica que más les gusta, sin embargo, la relación es baja.

¿Qué técnica detecta más faltas? (IT1) / ¿Qué técnica te ha parecido más fácil de entender? (PT2)

En ninguna de las tres iteraciones, hay acuerdo, por lo tanto, no existe tendencia a contestar lo mismo. Se puede concluir que los sujetos no identifican la técnica más efectiva con la técnica que les ha parecido más fácil de entender.

¿Qué técnica detecta más faltas? (IT1) / ¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

Antes de aplicar las técnicas, en ninguna de las tres iteraciones, hay acuerdo, por lo tanto, no existe tendencia a contestar lo mismo. Se puede concluir que parece que antes los sujetos no identifican la técnica más efectiva con la técnica que les ha parecido más fácil de aplicar.

Después aplicar las técnicas, en la iteración 1, no hay acuerdo, sin embargo, en la iteración 2 y 3, hay un acuerdo discreto, por lo tanto, parece que existe tendencia a contestar lo mismo. Debido a que en dos iteraciones hay acuerdo (aunque es bajo) y solo en una no hay acuerdo, se puede concluir que, debido a que los sujetos cambian de opinión, parece que después los sujetos identifican la técnica más efectiva con la técnica que les ha parecido más fácil de aplicar, sin embargo, la relación es baja.

¿Qué técnica has aplicado mejor? (IT2) / ¿Qué técnica te gusta más? (PT1)

En la iteración 1, no hay acuerdo, sin embargo, en la iteración 2 hay un acuerdo sustancial y en la iteración 3 hay un acuerdo moderado, por lo tanto, existe tendencia a contestar lo mismo. Debido a que en dos iteraciones hay bastante acuerdo y solo en una no hay acuerdo, se puede concluir que, los sujetos identifican la técnica que han aplicado mejor con la técnica que les gusta más, hay que tener en cuenta que la relación es intermedia.

¿Qué técnica has aplicado mejor? (IT2) / ¿Qué técnica te ha parecido más fácil de entender? (PT2)

En ninguna de las tres iteraciones, existe acuerdo, por lo tanto, no existe tendencia a contestar lo mismo. Se puede concluir que los sujetos no identifican la técnica que han aplicado mejor con la técnica que les ha parecido más fácil de entender.

¿Qué técnica has aplicado mejor? (IT2) / ¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

En las iteraciones 1 y 3, hay un acuerdo moderado, además en la iteración 2 el acuerdo es sustancial, por lo tanto, existe tendencia a contestar lo mismo. Se puede concluir que los sujetos identifican la técnica que han aplicado mejor con la técnica que les ha parecido más fácil de aplicar, hay que tener en cuenta que la relación es intermedia.

8.3.2. Resultados sobre los Programas

¿Con qué programa has detectado más faltas? (IP1) / ¿Qué programa es más fácil? (PP1)

En la iteración 2, no hay acuerdo, además, en la iteración 3, hay un acuerdo discreto, por lo tanto, parece que no existe tendencia a contestar lo mismo. Debido a que en una iteración no hay acuerdo y en la otra hay poco acuerdo, se puede concluir que los sujetos no identifican el programa en el que han detectado más faltas con el programa más fácil.

¿Con qué programa has detectado más faltas? (IP1) / ¿Qué programa tiene las faltas más difíciles? (PP2)

En las dos iteraciones hay un desacuerdo discreto, por lo tanto, no existe tendencia a contestar lo mismo. Se puede concluir que, como era de esperar, los sujetos no identifican el programa en el que han detectado más faltas con el programa que tiene las faltas más difíciles.

¿Con qué programa has detectado más faltas? (IP1) / ¿Qué programa has entendido mejor? (PP3)

En la iteración 3 hay un acuerdo discreto, por lo tanto, parece que existe cierta tendencia a contestar lo mismo. Se puede concluir que, los sujetos identifican el programa en el que han detectado más faltas con el programa que han entendido mejor. Hay que tener en cuenta que la relación es baja.

¿Con qué programa has detectado más faltas? (IP1) / ¿Qué programa tiene las faltas más fáciles? (PP4)

En la iteración 3 hay un acuerdo discreto, por lo tanto, parece que existe cierta tendencia a contestar lo mismo. Se puede concluir que, los sujetos identifican el programa en el que han detectado más faltas con el programa que tiene las faltas más fáciles. Hay que tener en cuenta que la relación es baja.

¿Con qué programa has detectado menos faltas? (IP2) / ¿Qué programa es más fácil? (PP1)

En la iteración 3 hay un desacuerdo discreto, por lo tanto, parece que no existe cierta tendencia a contestar lo mismo. Se puede concluir que, tal como era de esperar, los sujetos no identifican el programa en el que han detectado menos faltas con el programa que les parece más fácil.

¿Con qué programa has detectado menos faltas? (IP2) / ¿Qué programa tiene las faltas más difíciles? (PP2)

En la iteración 3, se observa que hay un acuerdo moderado, por lo tanto, existe cierta tendencia a contestar lo mismo. Se puede concluir que, los sujetos identifican el programa en el que han detectado menos faltas con el programa que tiene las faltas más difíciles. Hay que tener en cuenta que la relación es intermedia.

¿Con qué programa has detectado menos faltas? (IP2) / ¿Qué programa has entendido mejor? (PP3)

En la iteración 3 hay un desacuerdo discreto, por lo tanto, parece que no existe cierta tendencia a contestar lo mismo. Se puede concluir que, tal como era de esperar, los sujetos no identifican el programa en el que han detectado menos faltas con el programa que han entendido mejor.

¿Con qué programa has detectado menos faltas? (IP2) / ¿Qué programa tiene las faltas más fáciles? (PP4)

En la iteración 3 hay un desacuerdo discreto, por lo tanto, no existe tendencia a contestar lo mismo. Se puede concluir que, tal como era de esperar, los sujetos no identifican el programa en el que han detectado menos faltas con el programa que tiene las faltas más fáciles.

8.3.3. Resultados Globales

En definitiva, con respecto a las técnicas, los sujetos no identifican, antes de aplicar las técnicas, la técnica más efectiva con su técnica preferida ni con la más fácil de aplicar y después de aplicar las técnicas, tampoco identifican la técnica más efectiva con la más fácil de entender. Sin embargo, después de aplicar las técnicas, identifican la técnica más efectiva con su técnica preferida y con la más fácil de aplicar, aunque la relación existente en general, es baja. También identifican la técnica que han aplicado mejor con su técnica preferida y con la que les ha parecido más fácil de aplicar, además la relación existente es intermedia.

Con respecto a los programas, los sujetos no identifican el programa más efectivo con el programa más fácil, ni con el programa que tiene las faltas más difíciles, tampoco identifican el programa menos efectivo con el programa más fácil, ni con el programa que han entendido mejor, ni tampoco con el programa que tiene las faltas más fáciles. Sin embargo, identifican el programa en el que han detectado más faltas con el programa que han entendido mejor, y con el que tiene las faltas más fáciles, aunque la relación existente es baja. Finalmente, también identifican el programa menos efectivo con el que tiene las faltas más difíciles, además la relación existente es intermedia.

8.4. Intuiciones vs Realidad

A continuación, se muestran todas las preguntas relacionadas con la intuición, que se han analizado a lo largo de este capítulo. Únicamente, se muestran los valores de la interacción de la técnica o programa elegido (intuición) con la técnica o programa aplicado (realidad). Para determinar en qué casos la intuición coincide con la Realidad. Todos los resultados se pueden contrastar en la Tabla 8.8.

8.4.1. Resultados sobre las Técnicas

¿Qué técnica detecta más faltas? (IT1)

Para la iteración 1, tanto en las respuestas antes del estudio empírico como en las de después, independientemente de lo que intuyan los sujetos, encuentran más faltas cuando aplican “Funcional”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando responden “Estructural”. Por lo tanto, habría intuición cuando los sujetos respondan “Funcional” y “Estructural”.

Para la iteración 2, tanto en las respuestas antes del estudio empírico como en las de después, y en la iteración 3, solo en las respuestas antes del estudio empírico, independientemente de lo que intuyan los sujetos, encuentran el mismo número de faltas con las tres técnicas. Por lo tanto, habría intuición cuando los sujetos respondan cualquiera de las tres técnicas.

Para la iteración 3, en las respuestas después del estudio empírico, cuando los sujetos responden “Revisión”, encuentran más faltas con dicha técnica, sin embargo, cuando responden

“Estructural” y “Funcional” encuentran el mismo número de faltas con las tres técnicas. Por lo tanto, siendo estrictos, solo habría intuición cuando responden “Revisión”, pero dado que solo se permitía elegir una respuesta, se puede aceptar intuición en los tres casos.

Pregunta		Iteración 1		Iteración 2		Iteración 3	
		Elección	Realidad	Elección	Realidad	Elección	Realidad
IT1	A	R,E,F	F>R	R,E,F	R=E=F	R,E,F	R=E=F
			R=E				
			E=F				
	D	R,E,F	F>E	R,E,F	R=E=F	R	R>(E=F)
			R=E			E	R=E=F
			E=F			F	
IT2	R	R	F>R	R>E	R,E,F	R>F	
			R=E	R=F			
			R=F	E=F		R=E	
	E,F	E F	F>R	R=E=F			E=F
			R=E				
			E=F				
IP1	No aparece en esta iteración			cm	cm=na=nt	cm,na,nt	na>cm
				na	na>cm		cm=nt
					cm=nt		na=nt
					na=nt		(na=nt)>cm
				nt			
IP2	No aparece en esta iteración			No aparece en esta iteración		cm	(na=nt)>cm
						na	cm=na=nt
						nt	(cm=nt)<na

Tabla 8.8. Resumen de Intuiciones vs Realidad

¿Qué técnica has aplicado mejor? (IT2)

En la Tabla 8.8, con las repuestas proporcionadas después de aplicar las técnicas, se observa que, para la iteración 1, cuando los sujetos intuyen “Revisión”, encuentran más faltas cuando aplican “Funcional”, es decir, la intuición es totalmente errónea. Cuando responden “Estructural” y “Funcional” encuentran más faltas cuando aplican “Funcional”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando respondan “Estructural”. Por lo tanto, solo habría intuición cuando los sujetos respondan “Funcional” y “Estructural”.

Para la iteración 2, cuando los sujetos intuyen “Revisión”, encuentran más faltas cuando aplican dicha técnica, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando respondan “Funcional”. Cuando responden “Estructural” y “Funcional” encuentran el mismo número de faltas con las tres técnicas. Por lo tanto, siendo estrictos, solo habría intuición cuando responden “Revisión”, pero dado que solo se permitía elegir una respuesta, se puede aceptar intuición en los tres casos.

Para la iteración 3, independientemente de lo que los sujetos intuyan, encuentran más faltas “Revisión”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando respondan “Estructural”. Por lo tanto, solo habría intuición cuando los sujetos respondan “Revisión” y “Estructural”.

8.4.2. Resultados sobre los Programas

¿Con qué programa has detectado más faltas? (IP1)

En la Tabla 8.8, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que, para la iteración 2, cuando los sujetos intuyen “Cmdline”, encuentran el mismo número de faltas en los tres programas. Cuando responden “Nametbl” encuentran más faltas en dicho programa, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando respondan “Ntree”. Finalmente, cuando responden “Ntree” encuentran más faltas en el mismo y en “Nametbl”. Por lo tanto, siendo estrictos solo habría intuición cuando los sujetos respondan “Nametbl” y “Ntree”, pero dado que solo se permitía una respuesta, se puede aceptar intuición en los tres casos.

Para la iteración 3, independientemente, de la intuición de los sujetos encuentran más faltas en “Nametbl”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que existe intuición cuando respondan “Ntree”. Por lo tanto, solo habría intuición cuando ellos responden “Nametbl” y “Ntree”.

¿Con qué programa has detectado menos faltas? (IP2)

En la Tabla 8.8, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que, para la iteración 3, cuando los sujetos intuyen “Cmdline”, encuentran menos faltas en este programa. Cuando responden “Nametbl” encuentran el mismo número de faltas en los tres programas. Finalmente, cuando responden “Ntree” encuentran menos faltas en el mismo y en “Cmdline”. Por lo tanto, siendo estrictos solo habría intuición cuando los sujetos respondan “Cmdline” y “Ntree”, pero dado que solo se permitía elegir una respuesta, se puede aceptar intuición en los tres casos.

8.4.3. Resultados Globales

En definitiva, para las preguntas relacionadas con las técnicas, para la técnica más efectiva, en las respuestas dadas antes y después de aplicar las técnicas, para las iteraciones 2 y 3 los sujetos aplican las tres técnicas igual, aunque siendo estrictos después de aplicar las técnicas en la iteración 3 solo aciertan los que responden “Revisión”, sin embargo en la iteración 1 aciertan los que responden “Funcional” y “Estructural”. Sin embargo, en las respuestas anteriores a la aplicación de las técnicas, no existe una intuición clara sobre la técnica más efectiva por lo tanto, no se puede considerar que exista intuición. La intuición después de aplicar las técnicas, en la iteración 3 son “Revisión” y “Funcional”, por lo tanto, en este caso se podría considerar que existe intuición sobre la técnica más efectiva. En general, parece que la intuición sobre la técnica más efectiva no sirve para predecir.

Para la técnica que han aplicado mejor para las iteraciones 1 y 2 los sujetos aplican las tres técnicas igual, sin embargo en la iteración 3 aciertan los que responden “Revisión” y “Estructural”. Y

las intuiciones de los sujetos sobre la técnica que han aplicado mejor en las iteraciones 1 y 2 es “Funcional”, por lo tanto, no habría coincidencia. Es decir, la intuición no sirve para predecir. En el caso de la iteración 3, ocurre lo mismo que en el caso anterior, las intuiciones sobre la técnica que han aplicado mejor son “Revisión” y “Funcional”. Hay que tener en cuenta que solo coincide una de las intuiciones con una de las técnicas que aplicaron mejor. Por lo tanto, la coincidencia no es rotunda. En general, parece que la intuición sobre la técnica que han aplicado mejor no sirve para predecir.

Para las preguntas relacionadas con los programas, para el programa más efectivo los sujetos que responden “Nmetbl” y “Ntree”, intuyen correctamente. Para el programa menos efectivo intuyen correctamente los sujetos que responden “Cmdline” y “Ntree”. Sin embargo, para ninguno de los dos casos existe una intuición definida, por lo tanto, no se puede considerar que exista intuición. Es decir, parece que la intuición no sirve para predecir.

8.5. Preferencias vs Realidad

A continuación, se muestran todas las preguntas relacionadas con las preferencias, que se han analizado a lo largo de este capítulo. Únicamente, se muestran los valores de la interacción de la técnica o programa elegido (preferencia) con la técnica o programa aplicado (realidad). Para determinar en qué casos las preferencias coinciden con la Realidad. Todos los resultados se pueden contrastar en la Tabla 8.9.

8.5.1. Resultados sobre las Técnicas

¿Qué técnica te gusta más? (PT1)

En la Tabla 8.6, con las repuestas proporcionadas antes de aplicar las técnicas, se observa que, para la iteración 1, cuando los sujetos responden “Revisión” encuentran el mismo número de faltas con las tres técnicas. Cuando responde “Estructural” encuentran más faltas cuando aplican “Funcional”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Estructural”. Finalmente, cuando responden “Funcional” encuentran más faltas cuando aplican esta técnica. Por lo tanto, siendo estrictos solo influirían las preferencias cuando los sujetos respondan “Funcional”, pero dado que solo se permitía elegir una respuesta, se puede aceptar que en los tres casos hay influencia.

Para la iteración 2, tanto en las respuestas antes del estudio empírico como en las de después, independientemente de las preferencias de los sujetos, encuentran el mismo número de faltas con las tres técnicas. Por lo tanto, las preferencias influirían cuando los sujetos respondan cualquiera de las tres técnicas.

Para la iteración 3, tanto en las respuestas antes del estudio empírico como en las de después, independientemente de las preferencias de los sujetos, encuentran más faltas cuando aplican “Revisión”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede

aceptar que las preferencias influyen cuando respondan “Estructural”. Por lo tanto, las preferencias influirían solo cuando los sujetos respondan “Revisión” y “Estructural”.

Con las repuestas proporcionadas después de aplicar las técnicas, se observa que, para la iteración 1, cuando los sujetos responden “Revisión” y “Estructural” encuentran el mismo número de faltas con las tres técnicas. Cuando responden “Funcional” encuentran más faltas cuando aplican esta técnica. Por lo tanto, siendo estrictos solo influirían las preferencias cuando los sujetos respondan “Funcional”, pero dado que solo se permitía elegir una respuesta, se puede aceptar que en los tres casos hay influencia.

		Iteración 1		Iteración 2		Iteración 3		
Pregunta		Elección	Realidad	Elección	Realidad	Elección	Realidad	
PT1	A	R	R=E=F	R,E,F	R=E=F	R,E,F	R>F	
		E	F>R				R=E	
			R=E				E=F	
			F				F>(R=E)	E=F
	D	R,E	R=E=F	R,E,F	R=E=F	R,E,F	R>F	
		F	F>(R=E)				R=E	
E=F								
PT2		R,E,F	F>R	R,E,F	R>E	R,E,F	R>F	
			R=E				R=F	R=E
			E=F				E=F	E=F
PT3	A	R,E	R=E=F	R,E,F	R=E=F	R,E,F	R>F	
		F	F>R				R=E	
			R=E				E=F	
			E=F				E=F	
	D	R,E,F	F>R	R,E,F	R=E=F	R,E,F	R>F	
			R=E				R=E	
E=F			E=F					
PP1		cm,na,nt	cm=na=nt	cm,na,nt	(na=nt)>cm	cm,na,nt	na>cm	
							cm=nt	
							na=nt	
PP2		cm,na,nt	cm=na=nt	cm,na,nt	(na=nt)>cm	cm	cm<na	
							cm=nt	
							na=nt	
						na	cm=na=nt	
						nt	(cm=nt)<na	
PP3		No aparece en esta iteración		cm,na,nt	(na=nt)>cm	cm,na,nt	(cm=nt)<na	
PP4		No aparece en esta iteración		No aparece en esta iteración		cm,na,nt	na>cm	
							cm=nt	
							na=nt	

Tabla 8.9. Resumen de Preferencias vs Realidad

¿Qué técnica te ha parecido más fácil de entender? (PT2)

En la Tabla 8.9, con las repuestas proporcionadas después de aplicar las técnicas, se observa que, para la iteración 1, independientemente de lo que los sujetos respondan encuentran más faltas en “Funcional”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Estructural”. Por lo tanto, solo influirían las preferencias cuando los sujetos respondan “Funcional” y “Estructural”.

Para la iteración 2, independientemente de las preferencias de los sujetos, encuentran más faltas cuando aplican “Revisión”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que las preferencias influyen cuando respondan “Funcional”. Por lo tanto, las preferencias influirían solo cuando los sujetos respondan “Revisión” y “Funcional”.

Para la iteración 3, independientemente de las preferencias de los sujetos, encuentran más faltas cuando aplican “Revisión”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que las preferencias influyen cuando respondan “Estructural”. Por lo tanto, las preferencias influirían solo cuando los sujetos respondan “Revisión” y “Estructural”.

¿Qué técnica te ha parecido más fácil de aplicar? (PT3)

En la Tabla 8.9, en las repuestas proporcionadas antes de aplicar las técnicas, en la iteración 1 cuando responden “Revisión” y “Estructural” encuentran el mismo número de faltas con las tres técnicas. Cuando responden “Funcional” encuentran más faltas en esta técnica, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Estructural”. Por lo tanto, siendo estrictos solo influirían las preferencias cuando los sujetos respondan “Funcional” y “Estructural”, pero dado que solo se permitía elegir una respuesta, se puede aceptar que en los tres casos hay influencia.

Para la iteración 2, tanto en las respuestas antes del estudio empírico como en las de después, independientemente de las preferencias de los sujetos, encuentran el mismo número de faltas con las tres técnicas. Por lo tanto, las preferencias influirían cuando los sujetos respondan cualquiera de las tres técnicas.

Para la iteración 3, tanto en las respuestas antes del estudio empírico como en las de después, independientemente de las preferencias de los sujetos, encuentran más faltas cuando aplican “Revisión”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que las preferencias influyen cuando respondan “Estructural”. Por lo tanto, las preferencias influirían solo cuando los sujetos respondan “Revisión” y “Estructural”.

Con las repuestas proporcionadas después de aplicar las técnicas, se observa que, para la iteración 1, independientemente de las respuestas de los sujetos, encuentran más faltas con “Funcional”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Estructural”. Por lo tanto, solo influirían las preferencias cuando los sujetos respondan “Funcional” y “Estructural”.

8.5.2. Resultados sobre los Programas

¿Qué programa es más fácil? (PP1)

En la Tabla 8.9, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que, para la iteración 1, independientemente de lo que los sujetos respondan, encuentran el mismo número de faltas en los tres programas. Por lo tanto, las preferencias influirían cuando los sujetos respondan cualquiera de los tres programas.

Para la iteración 2, independientemente de lo que los sujetos respondan, encuentran más faltas en los programas “Nametbl” y “Ntree”. Por lo tanto, las preferencias influirían cuando los sujetos respondan “Nametbl” y “Ntree”.

Para la iteración 3, independientemente de lo que los sujetos respondan, encuentran más faltas en “Nametbl”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Ntree”. Por lo tanto, las preferencias influirían cuando los sujetos respondan “Nametbl” y “Ntree”.

¿Qué programa tiene las faltas más difíciles? (PP2)

En la Tabla 8.9, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que, para la iteración 1, independientemente de lo que los sujetos respondan, encuentran el mismo número de faltas en los tres programas. Por lo tanto, las preferencias influirían cuando los sujetos respondan cualquiera de los tres programas.

Para la iteración 2, independientemente de lo que los sujetos respondan, encuentran menos faltas en “Cmdline”. Por lo tanto, las preferencias influirían solo cuando los sujetos respondan “Cmdline”.

Para la iteración 3, cuando los sujetos responden “Cmdline” encuentran menos faltas en este programa, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Ntree”. Cuando responden “Nametbl” encuentran el mismo número de faltas en los tres programas. Finalmente, cuando responden “Ntree” encuentran menos faltas en este programa y en “Ntree”. Por lo tanto, siendo estrictos las preferencias influirían cuando los sujetos respondan “Cmdline” y “Ntree”, pero dado que solo se permitía elegir una respuesta, se puede aceptar que en los tres casos hay influencia.

¿Qué programa has entendido mejor? (PP3)

En la Tabla 8.9, con las repuestas proporcionadas después de aplicar las técnicas sobre los programas, se observa que, para la iteración 2, independientemente de lo que los sujetos respondan, encuentran más faltas en los programas “Nametbl” y “Ntree”. Por lo tanto, las preferencias influirían cuando los sujetos respondan “Nametbl” y “Ntree”.

Para la iteración 3, independientemente de lo que los sujetos respondan encuentran más faltas en “Nametbl”. Por lo tanto, las preferencias influirían cuando los sujetos respondan “Nametbl”.

¿Qué programa tiene las faltas más fáciles? (PP4)

En la Tabla 8.9, con las repuestas proporcionadas después de aplicar las técnicas sobre cada programa, se puede observar que, independientemente de lo que los sujetos respondan, en la iteración 3, encuentran menos faltas en “Cmdline”, teniendo en cuenta que solo se les permitía elegir una respuesta, también se puede aceptar que la preferencia influye cuando respondan “Ntree”. Por lo tanto, las preferencias influirían únicamente cuando los sujetos respondan “Cmdline” y “Ntree”.

8.5.3. Resultados Globales

En definitiva, para las preguntas relacionadas con las técnicas, para la técnica preferida, en las respuestas dadas antes y después de aplicar las técnicas, en la iteración 2 se aplican las tres técnicas igual. En la iteración 3 aciertan los que responden “Revisión” y “Estructural”, sin embargo antes de aplicar las técnicas no hay una preferencia clara y después de aplicarlas, las técnicas preferida de los sujetos son “Funcional” y “Revisión”, es decir, solo coinciden con “Funcional”, es decir hay poca coincidencia entre la técnica preferida y la que se ha aplicado mejor. Finalmente, para la iteración 1, después de aplicar las técnicas aciertan los que responden “Funcional” y antes de aplicar las técnicas también aciertan los que responden “Estructural”. Sin embargo, solo se podría considerar acuerdo antes de aplicar las técnicas, para las dos técnicas porque son las preferidas de los sujetos, tanto antes como después de aplicar las técnicas, en el caso del después solo coinciden los de “Funcional”, es decir, hay poca coincidencia entre la técnica preferida y la que se ha aplicado mejor. Por lo tanto, parece que en general las preferencias influyen poco sobre la aplicación de las técnicas.

Para la técnica más fácil de entender, en la iteración 1, coinciden los sujetos que responden “Funcional” y “Estructural”, sin embargo como no hay una preferencia clara, se puede considerar que no está relacionada la técnica más fácil de entender con la que han aplicado mejor. En la iteración 2, coinciden los sujetos que responden “Revisión” y “Funcional”, además los sujetos responden precisamente estas técnicas como las más fáciles de entender, se puede considerar que está relacionada la técnica más fácil de entender con la que han aplicado mejor. Finalmente, para la iteración 3, coinciden los sujetos que responden “Revisión” y “Estructural”, sin embargo como no hay una preferencia clara, se puede considerar que no está relacionada la técnica más fácil de entender con la que han aplicado mejor. En general, parece que las preferencias no sirven para predecir.

Para la técnica más fácil de aplicar, en la iteración 1, en las respuestas de antes, coinciden los sujetos que responden “Funcional” y “Estructural”, sin embargo como no hay una preferencia clara, se puede considerar que no está relacionada la técnica más fácil de entender con la que han aplicado mejor. En las preguntas de después, coinciden los sujetos que responden “Funcional” y “Estructural” y las preferencias de los sujetos sobre la técnica más fácil de aplicar es “Funcional”, se puede considerar que en cierta manera están relacionadas. En la iteración 2, aplican las tres técnicas por igual. Finalmente, para la iteración 3, coinciden los sujetos que responden “Revisión” y “Estructural”, sin embargo la preferencia de los sujetos es “Funcional”, se puede considerar que no está relacionada la técnica más fácil de entender con la que han aplicado mejor. En general, parece que las preferencias no sirven para predecir.

Para la técnica que han aplicado mejor para las iteraciones 1 y 2 los sujetos aplican las tres técnicas igual, sin embargo en la iteración 3 aciertan los que responden “Revisión” y “Estructural”. Y las intuiciones de los sujetos sobre la técnica que han aplicado mejor en las iteraciones 1 y 2 es “Funcional”, por lo tanto, no habría coincidencia. Es decir, la intuición no sirve para predecir. En el caso de la iteración 3, ocurre lo mismo que en el caso anterior, las intuiciones sobre la técnica que han aplicado mejor son “Revisión” y “Funcional”. Hay que tener en cuenta que solo coincide una de

las intuiciones con una de las técnicas que aplicaron mejor. Por lo tanto, la coincidencia no es rotunda. En general, la intuición sobre la técnica que han aplicado mejor no sirve para predecir.

Para las preguntas relacionadas con los programas, para el programa más fácil, el mejor entendido y el que tiene las faltas más fáciles los sujetos que responden “Nmetbl” intuyen correctamente. Para el programa que tiene las faltas más difíciles intuyen correctamente los sujetos que responden “Cmdline” y “Ntree”. Sin embargo, para ninguno de los cuatro casos existe una preferencia definida, por lo tanto, no se puede considerar que las preferencias influyan en la aplicación de las técnicas sobre los programas. Es decir, las preferencias no sirven para predecir.

Es importante notar, que tanto en intuiciones como en preferencias, la efectividad el porcentaje de faltas encontradas coinciden entre las iteraciones y entre las distintas preguntas. Sin embargo, con las técnicas se obtienen resultados bastante diferentes dependiendo de la técnica elegida.

CAPÍTULO 9. CONCLUSIONES

En este capítulo, se presentan las conclusiones obtenidas después de llevar a cabo este Trabajo Fin de Máster, para ello se analiza la consecución de los objetivos planteados al inicio de este trabajo en la sección 11.2.

En dicha sección se partió de un objetivo general que se subdividió en 4 subobjetivos. El objetivo principal fue:

Estudiar si las intuiciones y/o preferencias de los sujetos al de aplicar tres técnicas de evaluación de código coinciden con la efectividad de las mismas

Para poder cumplir el objetivo general, adicionalmente, se plantearon los siguientes subobjetivos:

Objetivo 1: *Conocer las intuiciones y las preferencias de los sujetos en lo relativo a las técnicas de evaluación estudiadas.*

Con respecto a las intuiciones, en el caso de las técnicas, antes y después de aplicar las técnicas, **no hay una intuición clara** con respecto a la técnica más efectiva, debido a que los sujetos intuyen que las tres técnicas son igual de efectivas. Con respecto a la técnica que los sujetos consideran que han aplicado mejor, coinciden en que es **Funcional**, además se obtienen resultados contradictorios para las técnicas “Revisión” y “Estructural”. Además, los sujetos tienden a **identificar** la técnica más efectiva con la técnica que han aplicado mejor.

Con respecto a las intuiciones, en el caso de los programas, los sujetos **no tienen una intuición clara** con respecto al programa más efectivo, ni tampoco con respecto al menos efectivo, debido a que los tres programas se distribuyen uniformemente. Además, los sujetos **no identifican** el programa más efectivo con el menos efectivo.

Con respecto a las preferencias, en el caso de las técnicas, antes de aplicar las técnicas, los sujetos **no tienen claro** que técnica les gusta más, debido a que por un lado las tres técnicas se distribuyen igual y por otro lado hay contradicciones entre “Funcional” y “Estructural”, sin embargo, después de aplicar las técnicas la técnica que más les gusta es **Funcional**. Con respecto a la técnica

más fácil de entender no hay ninguna elección clara. Finalmente, la técnica que los sujetos consideran más fácil de aplicar es **Funcional**, tanto antes como después de aplicar las técnicas. Además, los sujetos tienden a **identificar** la técnica preferida con la más fácil de aplicar, tanto antes como después de aplicar las técnicas. Sin embargo, los sujetos **no identifican** la técnica preferida con la técnica más fácil de aplicar ni tampoco la técnica más fácil de entender con la más fácil de aplicar.

Con respecto a las preferencias, en el caso de los programas, los sujetos **no prefieren ningún programa** en concreto, como el programa que más fácil, ni como el que tiene las faltas más difíciles, ni como el que han entendido mejor y finalmente tampoco como el que tiene las faltas más fáciles, debido a que los tres programas se distribuyen uniformemente. Además, los sujetos tienden a **identificar** el programa más fácil con el que han entendido mejor y con el que tiene las faltas más fáciles y también el programa que han entendido mejor con el que tiene las faltas más fáciles. Sin embargo, los sujetos **no identifican** el programa más fácil con el que tiene las faltas más difíciles, no tampoco el programa con las faltas más difíciles con el que han entendido mejor y con el que tiene las faltas más fáciles.

Objetivo 2: *Averiguar si el hecho de aplicar las técnicas produce un cambio de opinión en los sujetos acerca de sus intuiciones y preferencias.*

Con respecto a las intuiciones, después de aplicar las técnicas, para la técnica más efectiva **existen cambios de opinión**, sin embargo, como se comentó, en el apartado anterior, este hecho no ha determinado *una intuición clara*.

Con respecto a las preferencias, después de aplicar las técnicas, con respecto a la técnica que más les gusta **existen cambios de opinión**, puesto que en los casos en los que hay acuerdo, el mismo es muy bajo, además como se comentó en el objetivo anterior este hecho ha determinado que la técnica que *más les gusta* a los sujetos sea “**Funcional**”. Con respecto a la técnica más fácil de aplicar también **existen cambios de opinión**, al igual que antes, en los casos en los que hay acuerdo, el mismo es demasiado alto, además este hecho ha determinado que la técnica que *más fácil de aplicar* sea **Funcional**.

Objetivo 3: *Averiguar si las intuiciones y las preferencias sirven para predecir la efectividad de las técnicas.*

Con respecto a las intuiciones, para la técnica más efectiva, antes de aplicar las técnicas, existen coincidencias para los sujetos cuya técnica más efectiva es **Funcional**. Después de aplicar las técnicas, existen coincidencias para los sujetos cuya técnica más efectiva es **Funcional y Revisión**. Para la técnica que han aplicado mejor existen coincidencias para los sujetos cuya técnica más efectiva es **Revisión** y también para los que responden **Funcional**. Sin embargo, estas coincidencias no son fiables, por lo tanto, *la intuición no sirve para predecir*.

Con respecto a las intuiciones, con respecto al programa más efectivo, existen coincidencias para los sujetos cuyo programa más efectivo es **Nametbl** y también para los que responden **Ntree**. Para el programa menos efectivo, existen coincidencias para los sujetos cuyo programa menos

efectivo es **Cmdline** y también para los que responden **Ntree**. Sin embargo, estas coincidencias no son fiables, por lo tanto, la **intuición no sirve para predecir**.

Con respecto a las preferencias, para la técnica preferida, tanto antes como después de aplicar las técnicas, existen coincidencias para los sujetos cuya técnica preferida es **Funcional**. Para la técnica más fácil de entender existen coincidencias para los sujetos cuya técnica más fácil de entender es **Funcional**. Finalmente para la técnica más fácil de aplicar, antes de aplicar las técnicas, existen coincidencias para los sujetos cuya técnica más fácil de aplicar es **Funcional**. Después de aplicar las técnicas, existen coincidencias para los sujetos cuya técnica más fácil de aplicar es **Funcional** y también para los que responden **Revisión**. Sin embargo, estas coincidencias no son fiables, por lo tanto, **las preferencias no sirven para predecir**.

Con respecto a las preferencias, en el caso de los programas, para el programa preferido, el que han entendido mejor y el que tiene las faltas más fáciles, existen coincidencias para los sujetos cuya respuesta es **Nametbl** y **Ntree**. Por otro lado, para el programa que los sujetos consideran tiene **las faltas más difíciles**, existen coincidencias para los sujetos que responden **Cmdline** y **Ntree**. Sin embargo, estas coincidencias no son fiables, por lo tanto, **las preferencias no sirven para predecir**.

Objetivo 4: *Estudiar si existe relación entre las intuiciones y las preferencias de los sujetos*

Después de aplicar las técnicas, los sujetos tienden a **identificar** la técnica más efectiva con su técnica preferida y la más fácil aplicar. Y también **identifican** la técnica que han aplicado mejor con su técnica preferida y la más fácil aplicar. Sin embargo, antes de aplicar las técnicas, los sujetos **no identifican** la técnica más efectiva con su técnica preferida, ni con la más fácil de entender, ni con la más fácil aplicar.

Los sujetos tienden a **identificar** programa más efectivo con el programa con el que han entendido mejor y el que tiene las faltas más fáciles. Y **también identifican** el programa menos efectivo con el que tiene las faltas más difíciles. Sin embargo, los sujetos **no identifican** el programa más efectivo con el programa más fácil ni con el que tiene las faltas más difíciles. Y **tampoco identifican** el programa menos efectivo con el programa más fácil, el que han entendido mejor y el que tiene las faltas más fáciles.

De los resultados obtenidos, se puede concluir que existen cambios de opinión y estos podrían estar relacionados con que después de realizar el estudio empírico los sujetos tienen una visión más acertada de las técnicas de evaluación de código.

La intuición tiene tres fases. Primero, el cerebro recopila datos de la experiencia; seguidamente los procesa de forma inconsciente y automática, y en tercer lugar aparece repentinamente el resultado o la conclusión de este procesamiento en la consciencia. Por tanto, dado que los sujetos tienen poca experiencia sobre la aplicación de técnicas de evaluación de código, en este caso la intuición no sirve para predecir.

Tanto la intuición como las preferencias, existen más coincidencias en las respuestas de después de aplicar las técnicas, por lo tanto, ambas funcionan mejor para deducir lo que ya ha pasado que para predecir.

Finalmente, respondiendo al objetivo general, las intuiciones y preferencias no parecen estar relacionadas con la efectividad de las técnicas de evaluación de código.

REFERENCIAS

- [1] Juristo N., Moreno A., Basics of Software Engineering Experimentation, Spain. 2001
- [2] Lázaro M., Juristo N., Marcos E., Do Testers' Preferences Have an Impact on Effectiveness? QUATIC 2010: 304-309
- [3] N. Juristo, S. Vegas, "Functional Testing, Structural Testing and Code Reading: What Fault Type do they each Detect? Empirical Methods and Studies", Software Engineering- Experiences from ESERNET. N. Springer-Verlag. 2785(12), pp. 235-261, 2003.
- [4] Aguilera del Pino A., Análisis de tablas de contingencia bidimensionales, Abril, 2005
- [5] Xuezheng Sun, Zhao Yang, Generalized McNemar's Test for Homogeneity of the Marginal Distributions, *SAS Global Forum*, 2008
- [6] Landis, JR and Koch, GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977; 33:159-174.
- [7] Marija Norušis. IBM SPSS Advanced Statistics 20. Prentice Hall